UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MAGISTRSKO DELO

ANALIZA VIZUALIZACIJ MASOVNIH PODATKOV, KI PRIKAZUJEJO POTI

Ljubljana, september 2017 MARUŠA TOSTOVRŠNIK

IZJAVA O AVTORSTVU

Podpisana Maruša Tostovršnik, študentka Ekonomske fakultete Univerze v Ljubljani, avtorica predloženega dela z naslovom Analiza vizualizacij masovnih podatkov, ki prikazujejo poti, pripravljenega v sodelovanju s svetovalko doc. dr. Mojco Bavdaž,

IZJAVLJAM

1. da sem predloženo delo pripravila samostojno;
2. da je tiskana oblika predloženega dela istovetna njegovi elektronski obliki;
3. da je besedilo predloženega dela jezikovno korektno in tehnično pripravljeno v skladu z Navodili za izdelavo zaključnih nalog Ekonomske fakultete Univerze v Ljubljani, kar pomeni, da sem poskrbela, da so dela in mnenja drugih avtorjev oziroma avtoric, ki jih uporabljam oziroma navajam v besedilu, citirana oziroma povzeta v skladu z Navodili za izdelavo zaključnih nalog Ekonomske fakultete Univerze v Ljubljani;
4. da se zavedam, da je plagiatorstvo – predstavljanje tujih del (v pisni ali grafični obliki) kot mojih lastnih – kaznivo po Kazenskem zakoniku Republike Slovenije;
5. da se zavedam posledic, ki bi jih na osnovi predloženega dela dokazano plagiatorstvo lahko predstavljalo za moj status na Ekonomski fakulteti Univerze v Ljubljani v skladu z relevantnim pravilnikom;
6. da sem pridobila vsa potrebna dovoljenja za uporabo podatkov in avtorskih del v predloženem delu in jih v njem jasno označila;
7. da sem pri pripravi predloženega dela ravnala v skladu z etičnimi načeli in, kjer je to potrebno, za raziskavo pridobila soglasje etične komisije;
8. da soglašam, da se elektronska oblika predloženega dela uporabi za preverjanje podobnosti vsebine z drugimi deli s programsko opremo za preverjanje podobnosti vsebine, ki je povezana s študijskim informacijskim sistemom članice;
9. da na Univerzo v Ljubljani neodplačno, neizključno, prostorsko in časovno neomejeno prenašam pravico shranitve predloženega dela v elektronski obliki, pravico reproduciranja ter pravico dajanja predloženega dela na voljo javnosti na svetovnem spletu preko Repozitorija Univerze v Ljubljani;
10. da hkrati z objavo predloženega dela dovoljujem objavo svojih osebnih podatkov, ki so navedeni v njem in v tej izjavi.

V Ljubljani, dne _____ Podpis študentke:_____

KAZALO

UVOD **1**

1 OPREDELITEV MASOVNIH PODATKOV **2**
1.1 Definicija masovnih podatkov 2
1.2 Področja uporabe masovnih podatkov 4
1.2.1 Zasebni sektor 5
1.2.2 Javni sektor 5
1.2.3 Praktični primeri uporabe masovnih podatkov 6
1.4 Metode in orodja masovnih podatkov 6
1.5 Izzivi masovnih podatkov 8
1.6 Prihodnost masovnih podatkov 9

2 VIZUALIZACIJA MASOVNIH PODATKOV **9**
2.1 Zgodovina vizualizacije podatkov 9
2.2 Vizualizacija masovnih podatkov 11
2.2.1 Vizualizacijske tehnike 12
2.2.2 Vizualizacijska orodja 13
2.3 Vodila in priporočila za dobro vizualizacijo podatkov 14
2.3.1 Načela dobre vizualizacije 14
2.3.2 Človekovo dojemanje vizualizacije 15
2.3.3 Pripovedovanje zgodbe skozi vizualizacijo 16
2.4 Prihodnost vizualizacij masovnih podatkov 17
2.5 Vizualizacija poti nekoč in danes 18

3 EMPIRIČNA RAZISKAVA O RAZUMLJIVOSTI VIZUALIZACIJ MASOVNIH PODATKOV **19**
3.1 Namen in cilji raziskave 19
3.2 Raziskovalne hipoteze 19
3.3 Metodologija raziskave 20
3.4 Potek empirične raziskave 22
3.4.1 Oblikovanje vprašalnika 22
3.4.2 Raziskovalni vzorec 23

4 ANALIZA REZULTATOV EMPIRIČNE RAZISKAVE **25**
4.1 Analiza rezultatov prve vizualizacije 25
4.1.1 Analiza rezultatov vizualizacijske skupine 26
4.1.2 Analiza rezultatov skupine s ključnimi besedami 31
4.1.3 Povzetek rezultatov prve vizualizacije 32
4.2 Analiza rezultatov druge vizualizacije 33
4.2.1 Analiza rezultatov vizualizacijske skupine 34

4.2.2 Analiza rezultatov skupine s ključnimi besedami 39
4.2.3 Povzetek rezultatov druge vizualizacije 40
4.3 Analiza rezultatov tretje vizualizacije 40
4.3.1 Analiza rezultatov vizualizacijske skupine 41
4.3.2 Analiza rezultatov skupine s ključnimi besedami 46
4.3.3 Povzetek rezultatov tretje vizualizacije 48
4.4 Primerjava prikazov in končne ugotovitve....48

SKLEP....51

LITERATURA IN VIRI....53

PRILOGE

KAZALO TABEL

Tabela 1: Razdelitev skupine udeležencev in tipa nalog.... 22
Tabela 2: Socialno demografske značilnosti (spol in starost) raziskovalnega vzorca 23
Tabela 3: Socialno demografske značilnosti (stopnja in smer izobrazbe) raziskovalnega vzorca 24
Tabela 4: Kodiranje udeležencev glede na socialno demografske značilnosti 24
Tabela 5: Rezultati poteka vizualiziranja prve vizualizacije vizualizacijske skupine.... 27
Tabela 6: Rezultati ujemanja udeležencevih ključnih besed z vnaprej postavljenim kriterijem pri prvi vizualizaciji.... 31
Tabela 7: Rezultati poteka vizualiziranja druge vizualizacije druge skupine 35
Tabela 8: Rezultati ujemanja udeležencevih ključnih besed z vnaprej postavljenim kriterijem pri drugi vizualizaciji.... 39
Tabela 9: Rezultati poteka vizualiziranja tretje vizualizacije vizualizacijske skupine 43
Tabela 10: Rezultati ujemanja udeležencevih ključnih besed z vnaprej postavljenim kriterijem pri drugi vizualizaciji.... 47
Tabela 11: Razvrstitev vizualizacij glede na njihovo razumljivost.... 48

KAZALO SLIK

Slika 1: Klasifikacija masovnih podatkov.... 3
Slika 2: Klasifikacija vizualizacijskih tehnik 12
Slika 3: Aktivnost na Twitterju po katastrofalnem tajfunu na Filipinih novembra 2013 ... 25

Slika 4: Prehajanje med blagovnimi znamkami mobilnih telefonov med nizozemskimi uporabniki 33
Slika 5: Prikaz logističnih poti s pomočjo brezžičnih radijskih valov 41

UVOD

Dandanes živimo v hitrem svetu, v katerem smo vseskozi priča ustvarjanju novih podatkov, in to na vsakem koraku. Izraz masovni podatki (angl. *big data*) se čedalje pogosteje pojavlja v javnosti. Prav ta pojav je začel v zadnjih letih zanimati mnogo raziskovalcev, ki ga vsak na svoj način poskušajo opredeliti in masovne podatke narediti uporabne.

Masovni podatki se nanašajo na ogromne količine podatkov, ki se pojavljajo v različnih oblikah, in to tako rekoč vsako sekundo v vseh sferah našega življenja. Da bi bolje razumeli masovne podatke in izkoristili njihov potencial, so se razvili različni pristopi in metode k obdelavi masovnih podatkov, med njimi tudi vizualizacijske tehnike.

Vizualizacija masovnih podatkov bi nam morala na jasen način prinesti vpogled v masovne podatke. Na lažji način bi nam morala prikazati vzorce in trende sicer kompleksnih podatkovij. Ali je res tako in kako blizu so ljudem izbrane vizualizacije masovnih podatkov, raziskujem v magistrski nalogi. V raziskavo vključene tri dejanske vizualizacije masovnih podatkov se med sabo razlikujejo glede na težavnost in način vizualizacije, vseeno pa imajo nekaj skupnega. Vsaka na svoj način prikazuje poti, tokove oziroma premike.

Namen raziskovalnega dela je na majhnem vzorcu raziskati, v kolikšni meri so vizualizacije masovnih podatkov razumljive in kako jasno podajajo svojo glavno temo prikaza, torej ali pripomorejo k boljšemu razumevanju sporočila vizualizacije ali pa vodijo do napačnega.

Predstaviti želim tudi napotke za nadaljnje raziskovanje, na primer kako vizualizacije poti na osnovi masovnih podatkov z vidika podajanja glavne teme prikaza narediti še razumljivejše.

Temeljne raziskovalne hipoteze so naslednje:

- vizualizacije masovnih podatkov gledamo skozi tradicionalne metode grafičnih predstavitev;
- vizualizacije masovnih podatkov pripomorejo k razumevanju bistvenega sporočila masovnih podatkov;
- pomanjkljiva vizualizacija masovnih podatkov lahko prinese napačno sporočilnost.

Za obravnavano temo magistrskega dela sem se odločila, ker se tudi sama na svoji poklicni poti vsakodnevno srečujem z masovnimi podatki in jih moram prikazati na razumljiv ter preprost način. Priča sem temu, kako masovni podatki resnično lahko pomagajo še izboljšati finančne napovedi in razumevanje potreb strank.

V magistrskem delu bom najprej uporabila deduktivni raziskovalni pristop, saj bo prvi del sestavljala teorija, kjer bom opredelila masovne podatke in njihovo vizualizacijo, v drugem,

empiričnem delu pa bom na osnovi treh dejanskih vizualizacij masovnih podatkov proučevala, kako dobro prikazujejo glavno temo prikaza. V teoretičnemu delu bom uporabljala sekundarne bibliografske podatke, ki jih bom pridobila večinoma v tuji strokovni literaturi. V empiričnem delu bom izvedla kombinacijo delno strukturiranega intervjuja in opazovalne študije, prevladovala bo slednja. Kombinacija teh dveh metod zbiranja primarnih podatkov mi bo omogočila pridobiti precej podrobne odgovore.

Magistrsko delo je namenjeno vsem, ki se zavedajo, da so podatki in vse, kar lahko izhaja iz njih, naša prihodnost, ter to opazijo že dandanes pri svojem delu. Delo je namenjeno tudi oblikovalcem vizualizacij masovnih podatkov, da lahko na podlagi rezultatov empiričnega dela še izboljšajo svoje vizualizacije.

V prvem poglavju magistrskega dela opredelim masovne podatke, opišem njihova področja uporabe in metode ter orodja za njihovo obdelavo. Prav tako poudarim, kakšni so izzivi masovnih podatkov in v čem je njihova prihodnost. V drugem poglavju nadaljujem z vizualizacijo masovnih podatkov. Opišem bogato zgodovino vizualizacije podatkov in tehnike ter orodja vizualizacije masovnih podatkov. Poseben poudarek namenim vodilom in priporočilom za dobro vizualizacijo. Opišem tudi, v čem je prihodnost vizualizacije masovnih podatkov in kakšna je bila vizualizacija poti nekoč ter kakšna je danes. Tretje poglavje je začetek empiričnega dela magistrskega dela. V njem podam namen in cilje raziskave, raziskovalne hipoteze, opišem metodologijo raziskave ter navsezadnje njen potek. V zadnjem, četrtem poglavju pa predstavim in analiziram rezultate, pridobljene z empirično raziskavo.

1 OPREDELITEV MASOVNIH PODATKOV

1.1 Definicija masovnih podatkov

Pri masovnih podatkih govorimo o podatkovjih, katerih velikost presega meje tipične podatkovne baze. So kot veliki bazeni podatkov, ki jih lahko hranimo, posredujemo naprej in analiziramo (Manyika et al., 2011). V strokovni literaturi se največkrat pojavi Gartnerjeva definicija masovnih podatkov, po kateri so masovni podatki informacije, ki jih zaznamujejo trije v-ji (Gartner, b.l.; Marr, 2014; Wang, 2012):

- **velikost** (angl. *volume)* je količina zbranih in obdelanih podatkov, ki z leti sunkovito narašča, predvsem zaradi novejše in zmogljivejše tehnologije;
- **hitrost** (angl. *velocity*) opisuje frekvenco, s katero so podatki izbrani, obdelani in analizirani ter posredovani naprej. Je pomemben kazalnik, saj nam pove, kako hitro podatki krožijo in pridejo v organizacijo oziroma na mesto, kjer se jih potrebuje. Na drugi strani pa ne smemo pozabiti, da je prav tako pomemben čas oziroma hitrost, ko se podatki, ki so na voljo, uporabijo pri sprejemanju odločitev;

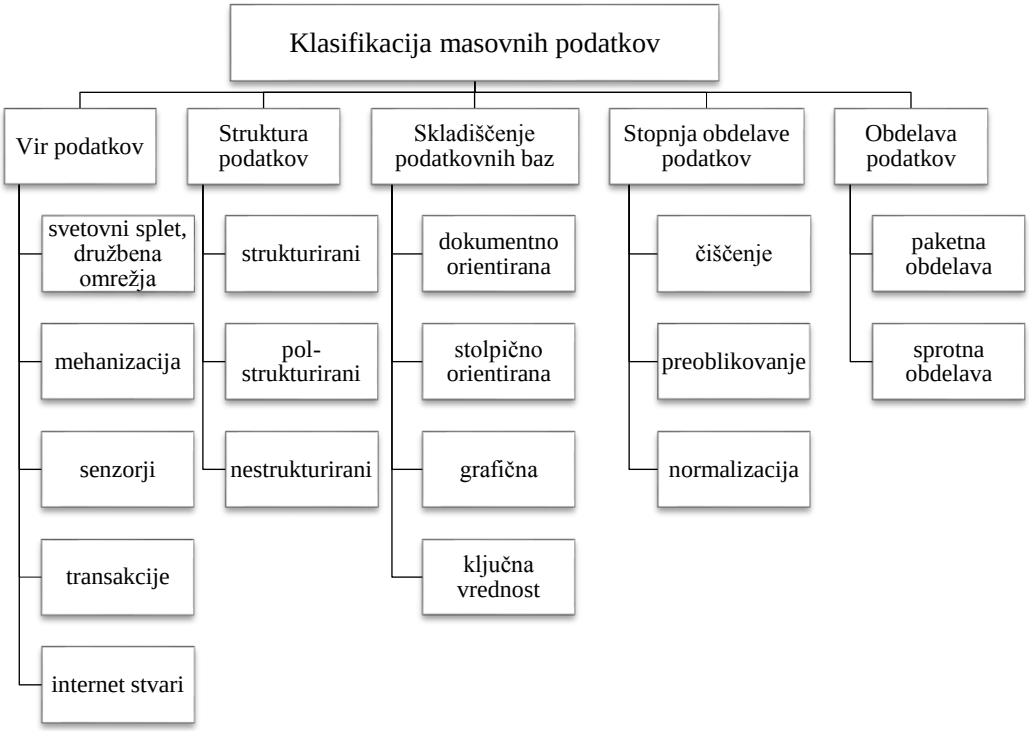
- **raznolikost** (angl. *variety*) zbranih podatkov se s časom močno spreminja, saj dandanes čedalje več podatkov prihaja iz različnih družbenih, mobilnih in drugih omrežij. Podatki spreminjajo tudi svojo strukturo, ki je bila v preteklosti veliko bolj definirana, kot je sedaj, ko je velika večina podatkov nestrukturiranih.

Poleg treh zgoraj omenjenih v-jev, ki se najpogosteje omenjajo v povezavi z masovnimi podatki, nekateri avtorji navajajo še naslednje (Desouza, 2014; Marr, 2014; Wang, 2012):

- **verodostojnost** (angl. *veracity*) meri pristranskost, »šum« in nepravilnosti podatkov. Velikokrat so namreč masovni podatki polni »šuma«, zato je pomembno, da ga prepoznamo in minimiziramo;
- **spremenljivost** (angl. *variability*) se nanaša na stopnjo spremenljivosti v pretoku podatkov;
- **nestalnost** (angl. *volatility*) nakazuje na to, kako dolgo so podatki veljavni in ali bi morali biti skladiščeni.

Masovne podatke lahko razvrstimo v različne skupine, da bolje razumemo njihove značilnosti. Klasifikacija temelji na petih vidikih, ki so prikazani na Sliki 1 (Hashem et al., 2015).

Slika 1: Klasifikacija masovnih podatkov



Vir: I. B. T. Hashem et al., The rise of "big data" on cloud computing: Review and open research issues, 2015, str. 101.

Vir podatkov (angl. *data source*) nakazuje, da se masovni podatki lahko pridobijo iz marsikatere platforme, od svetovnega spleta do podatkov, ki jih generirajo stroji v proizvodnih obratih, in interneta stvari (angl. *Internet of Things*) (Hashem et al. 2015).

Strukturirani masovni podatki so številke in besede, ki jih lahko preprosto razporedimo in analiziramo (Jayabharathi, 2015) v želenem časovnem okviru (Xu & Shi, 2015). Nestrukturirani podatki predstavljajo 95 odstotkov masovnih podatkov in so dostopni v avdio obliki, kot slike, posnetki in nestrukturiran tekst (Gandomi & Haider, 2015). Osnovna značilnost teh podatkov je, da so raznolični in izvirajo iz več različnih virov (Xu & Shi, 2015).

Dokumentno orientirana (angl. *document oriented*) skladišča podatkov so v večini namenjena skladiščenju in pridobivanju podatkov iz dokumentov. Stolpično orientirane (angl. *column-oriented*) podatkovne baze skladiščijo vsebino v stolpcih (Hashem et al., 2015). Grafične podatkovne baze (angl. *graph database*) so ustvarjene tako, da skladiščijo in predstavljajo podatke, ki uporabljajo grafične modele (Hashem et al., 2015). Metoda ključne vrednosti (angl. *key-value*) je alternativa relacijski podatkovni bazi, ko govorimo o skladiščenju in dostopnosti podatkov (Seeger, 2009).

Stopnje obdelave podatkov (angl. *data staging*) vključujejo čiščenje (angl. *cleaning*), ki je postopek prepoznavanja nepopolnih in nesmiselnih podatkov, preoblikovanje (angl. *transformation*), ki je proces preoblikovanja podatkov v primerne oblike za nadaljnjo analizo, in normalizacijo (angl. *normalization*) metod strukturiranja podatkovne baze za minimiziranje odvečnih podatkov (Hashem et al., 2015; Quackenbush, 2002; Rahm & Do, 2000).

Paketna obdelava podatkov (angl. *batch data processing*) je zelo učinkovit način obdelave podatkov. Podatki se najprej zberejo, vnesejo, obdelajo, šele nato se podajo rezultati paketov podatkov. Sprotna obdelava podatkov (angl. *real time data processing*) pa vključuje neprestane vnose, obdelave in rezultate podatkov (Walker, 2013).

1.2 Področja uporabe masovnih podatkov

Masovni podatki se uporabljajo na veliko področjih, od gospodarstva do javnega sektorja, od znanstvenih disciplin do nacionalne varnosti (Ask, Magnusson, & Bredmar, 2016). Eden izmed pomembnejših načinov uporabe masovnih podatkov je odkrivanje novih povezav in odvisnosti v podatkih, ki vodijo do gospodarskih priložnosti in večje učinkovitosti (Becker, 2016). Pomembnost masovnih podatkov je tudi v tem, da izboljšajo učinkovitost uporabe velikih baz podatkov, skladiščenje in upravljanje nestrukturiranih podatkov (Lokhande & Khare, 2015).

1.2.1 Zasebni sektor

Hkrati s količino informacij, ki je iz dneva v dan večja, se izboljšujejo tudi tehnologije za njihovo procesiranje. Nekatera podjetja so zaznala skokovit porast masovnih podatkov, ki jih želijo spremeniti v konkurenčno prednost. Na osnovi uporabe kvantitativnih analiz in prikaza teh podatkov lahko podjetja pravilno reagirajo in sprejemajo odločitve (Albright, Winston & Zappe, 2009).

Obstaja pet načinov, kako lahko masovni podatki ustvarjajo dodano vrednost za podjetja (Manyika et al., 2011):

- ustvarjanje **transparentnosti podatkov** oziroma omogočanje preprostega dostopa v potrebnem časovnem okviru;
- v današnjem času podjetja ustvarijo in shranijo večino podatkov v digitalni obliki, kar jim omogoča, da lahko podatke s pridom uporabijo za **napovedovanje poslovnih trendov**, temu pa prilagodijo vzvode in modele svojega poslovanja;
- masovni podatki omogočajo **natančno segmentacijo strank** in posledično tudi natančno prilagojene produkte in storitve;
- avtomatizirani algoritmi lahko bistveno **izboljšajo sprejemanje odločitev**;
- masovni podatki lahko pripomorejo k **izboljšanju razvoja novih izdelkov**.

Vse našteto se lahko uporabi v različnih vejah industrije – od razvoja, proizvodnje, prodaje, trženja do finančne industrije.

1.2.2 Javni sektor

Tudi javni sektor ustvari in zbere ogromne količine podatkov s svojimi vsakodnevnimi aktivnostmi, od upravljanja s pokojninami in davki, nacionalnimi sistemi zdravja do spremljanja prometa (Munné, 2016). Tudi za javni sektor je pomembna izraba potenciala tehnologije masovnih podatkov (Klievink, Romijn, Cunningham & de Bruijn, 2017). Pomagajo lahko pri sprejemanju odločitev in pri nadaljnjem razvoju omrežja javnih storitev (Milakovich, 2012). V javnem sektorju se lahko masovni podatki uporabljajo na številnih področjih, kot so zdravstvo, administracija, javni promet, vreme itd.

Uporaba masovnih podatkov v zdravstvu lahko izboljša zdravje posameznika in celotne populacije, strokovno učinkovitost, vodi pa lahko tudi do novih načinov zdravljenja nekaterih bolezni (Marconi, Dobra, & Thompson, 2013).

Zbiranje in analiza podatkov tako iz tradicionalnih kot tudi sodobnih družbenih omrežij in medijev lahko vladnim administracijam pomaga pri oblikovanju kakovostnejših storitev za svoje državljane. Tudi razni uradi, npr. davčni, so vedno bolj odvisni od uporabe masovnih

podatkov. Z vključitvijo strukturiranih in nestrukturiranih podatkov iz družbenih in drugih omrežij v svoje analize lahko potrdijo izračune ter hitreje odkrijejo zlorabe (Munné, 2016).

1.2.3 Praktični primeri uporabe masovnih podatkov

Poleg navedenih področij uporabe masovnih podatkov v nadaljevanju na kratko predstavljam nekaj praktičnih primerov njihove uporabe. Na osnovi opisanih primerov lahko dejansko vidimo, kako deluje uporaba masovnih podatkov.

Kosaraju (2015) navaja naslednja primera:

- **Walmart**, največji trgovec na drobno na svetu, na osnovi masovnih podatkov uporablja analize, s katerimi določi pozicioniranje izdelkov na prodajnih policah, ki služi povečanju prodaje;
- **Amazon,** največja spletna trgovina, na osnovi nakupovalnih navad svojih uporabnikov na spletu ustvari algoritme za sistem priporočil, s pomočjo katerih nato pošiljajo specifične ponudbe vsaki stranki.

Poleg zgoraj navedenih konkretnih primerov uporabe masovnih podatkov v zasebnem sektorju lahko navedemo še nekaj primerov njihove uporabe v javnem sektorju (Mangelsdorf, 2012; Marr, 2015):

- **NASA – Center za klimatske simulacije** (angl. *NCCS – Center for Climate Simulation*) ima vsakodnevno opravka z masovnimi podatki, saj v njihove superračunalnike vseskozi pritekajo podatki o klimatskih podatkih z vsega sveta. Strokovnjaki nato s pomočjo njihove vizualizacije simulirajo klimatske modele, s katerimi poizkušajo predvideti vremenske in atmosferske spremembe;
- **John Deere** je med drugim eden večjih proizvajalcev kmetijske mehanizacije na svetu. V zadnjih letih je razvil številne storitve, ki temeljijo na masovnih podatkih. Ena izmed njih je spletni portal, ki kmetom omogoča dostop do podatkov, zbranih prek senzorjev in dodanih k njihovi kmetijski mehanizaciji. Prav tako pa lahko dostopajo do podatkov ostali uporabniki po svetu. Ti podatki kmetom omogočajo učinkovitejšo uporabo njihove mehanizacije, poleg tega jim zagotavljajo tudi informacije o njenih morebitnih okvarah. Uporaba masovnih podatkov lahko kmetom pomaga tudi pri odločitvah, kje, kdaj in kaj posaditi, da bi dosegli maksimalne hektarske donose.

1.4 Metode in orodja masovnih podatkov

Tehnike obdelave masovnih podatkov vključujejo vrsto različnih disciplin, v katerih se tehnike med seboj prepletajo (Chen & Zhang, 2014).

V nadaljevanju navajam tehnike, ki se uporabljajo najpogosteje (Ali et al., 2016; Chen & Zhang, 2014; Rodríguez-Mazahua et al., 2016; Olshannikova, Ometov, Koucheryavy, & Olsson, 2015; Witten & Frank, 2005):

- **analiza družbenih omrežij** (angl. *social network analysis, SNA*) se je razvila kot ena ključnih metod v sodobni sociologiji in se lahko uporablja za analizo strukture medsebojnih razmerij;
- **napovedna analitika** (angl. *predictive analytics*) lahko prispeva k večji dodani vrednosti na način, da napoveduje prihodnje dogodke in obnašanja, ki temeljijo na preteklih izkušnjah. Vse torej temelji na tem, da se prihodnost lahko napove na osnovi zgodovinskih podatkov;
- **optimizacijske metode** (angl. *optimization methods*) so matematična orodja za učinkovito analizo podatkov. Optimizacija se osredotoča na iskanje optimalnega niza postopkov, potrebnih za izboljšanje delovanja sistema;
- **podatkovno rudarjenje** (angl. *data mining*) raziskuje in analizira velike količine podatkov, da bi se v njih odkrili smiselni vzorci, ki bi nam pomagali razumeti podatke, in na njihovi osnovi izdelati napovedi;
- **statistične tehnike** (angl. *statistics*) se uporabljajo za zbiranje, urejanje in interpretacijo podatkov ter tudi za poudarjanje medsebojnih povezav med proučevanimi objekti;
- **strojno učenje** (angl. *machine learning*) je pomembno področje umetne inteligence, njegov cilj pa je, da se računalniški sistemi na osnovi podatkov naučijo avtomatično izvajati določene naloge (npr. odločitve, napovedovanje);
- **vizualizacijske tehnike** (angl. *visualization techniques*) se nanašajo na vizualizacijo analitičnih rezultatov v obliki diagramov, tabel in slik, pri čemer se vizualizacije masovnih podatkov razlikujejo od tradicionalnih vizualizacijskih metod.

Masovni podatki zahtevajo tudi posebna orodja, da lahko v zahtevanem časovnem obdobju učinkovito procesirajo ogromne podatkovne baze (Rodríguez-Mazahua et al., 2016). Orodja masovnih podatkov lahko razdelimo v tri skupine (Chen & Zhang, 2014; Rodríguez-Mazahua et al., 2016):

- **orodja, s katerimi se podatki najprej shranijo in nato analizirajo** (angl. *Big Data tools based on batch analysis*), na primer: Google MapReduce, Apache Hadoop, Dryad;
- **orodja, ki analizirajo podatke, takoj ko je to mogoče** (angl. B*ig Data tools for stream analysis*), na primer: Apache Storm, Apache S4, Apache Spark, SAP Hana;
- **orodja za interaktivno analizo** (angl. *Big Data tools based on interactive analysis*), na primer: Apache Drill, SpagoBI.

1.5 Izzivi masovnih podatkov

Ustvarjanje in raziskovanje sistema masovnih podatkov je zapleteno področje. Poleg prednosti masovnih podatkov obstajajo tudi izzivi, ki jih lahko razdelimo v tri skupine (Lokhande & Khare, 2015):

- pridobivanje in upravljanje podatkov (angl. *data acquisition and management*);
- analitika podatkov (angl. *data analytics*);
- sistemski izzivi (angl. *system issues*).

Izzivi, povezani s pridobivanjem in z upravljanjem podatkov, vključujejo odpravo odvečnih ter nesmiselnih podatkov, stiskanje podatkov (angl. *data compression*), zasebnost in varnost podatkov (Lokhande & Khare, 2015), obstajajo pa tudi izzivi potrebnih kadrov za upravljanje z masovnimi podatki. V naboru podatkov se nahaja veliko število odvečnih in nesmiselnih podatkov. Zato je treba izvesti določene operacije za učinkovito odpravljanje teh pomanjkljivosti, ne da bi izgubili potencialno vrednost teh podatkov (Hu, Wen, Chua, & Li, 2014; Lokhande & Khare, 2015). Eden večjih in pomembnejših izzivov, pred katerimi so masovni podatki, je ravnotežje med zasebnostjo, varnostjo podatkov in ustvarjenimi podatki (Dean, 2014). Resnejši izziv zasebnosti je identifikacija zasebnih podatkov med prenosom na svetovni splet (Porambge et al., 2016). Da lahko resnično izkoristimo potencial masovnih podatkov, se potrebuje tudi pravšnji profil zaposlenih, da analizira te masovne podatke, a je vse prej kot lahko najti takšen kader (Kim, Trimi, & Chung, 2014).

Podatki, ki bodo osnova nekih odločitev, morajo biti primerno analizirani (Ularu, Puican, Apostu, & Velicanu, 2012). Ker nabor podatkov narašča in potrebe po sprotni obdelavi ter analizi podatkov postajajo čedalje pomembnejše, analiza celotnega nabora podatkov postaja velik izziv (Lokhande & Khare, 2015).

Med sistemske izzive, povezane z masovnimi podatki, vključujemo razširljivost oziroma stopnjevanost (angl. s*calability*) in nove tehnologije (angl. *new technologies*) (Lokhande & Khare, 2015). Količina podatkov še vedno eksponentno raste, zato morajo biti podjetja, ki želijo ostati dobičkonosna in konkurenčna, sposobna upravljati z vedno večjimi potrebami po skladiščenju in obdelavi masovnih podatkov. Vse to pomeni, da morajo biti njihovi sistemi masovnih podatkov nujno lahko in hitro razširljivi (Amster, 2016). Glede na to, da so masovni podatki relativno nov pojav, se morajo vsi, ki jih uporabljajo, naučiti, kako uporabljati nove tehnologije, povezane z njimi, takoj ko se pojavijo na trgu (Ularu et al., 2012).

1.6 Prihodnost masovnih podatkov

Pomembnost masovnih podatkov se bo nadaljevala in rasla tudi v prihodnosti. Količina podatkov, ki jih ustvarimo, eksponentno narašča, saj digitalni sistemi postajajo cenejši, aplikacije pa postajajo čedalje bolj povezane (Anagnostopoulos, Zeadally, & Exposito, 2016).

Masovni podatki ne bodo imeli samo socialnega in gospodarskega učinka, ampak bodo vplivali tudi na to, kako posamezniki živimo in razmišljamo (Chen, Mao, & Liu, 2014). Postali bodo novo sredstvo podjetja in pomembna konkurenčna prednost (Brown, Chui, & Manyika, 2011).

Da bi v prihodnosti bili kos obdelavi čedalje večje količine podatkov v najhitrejšem možnem času, se bo še naprej morala razvijati odprtokodna tehnologija (angl. *Open source technology*), saj jo bodo rabili novi programski jeziki za zadoščanje potreb (Chaudhari & Patel, 2017). Razvoj napovedne analitike oziroma razvoj modeliranja novih napovednih algoritmov bo še naprej strmo naraščal, tako da bo mogoče napovedne algoritme še bolje uporabiti v posebnih situacijah kot do sedaj (Dean, 2014).

Izpostaviti je treba tudi čedalje hitrejši razvoj interneta stvari, ki je kombinacija fizičnih in digitalnih komponent, s pomočjo katerih ustvarja nove proizvode in prispeva k novim poslovnim modelom (Wortmann & Flüchter, 2015). V prihodnosti bo čedalje več naprav komuniciralo med sabo namesto nas (Dean, 2014). Samo pomislimo, da že danes obstaja veliko pametnih domov, pametnih mest, v prihodnosti pa se bo to samo še nadgrajevalo. Obeti in pričakovanja glede interneta stvari so veliki, hkrati pa obstaja še kar precej izzivov, tako tehnoloških kot poslovnih. Vpeljava povezanih naprav v poslovanje namreč vzbudi nekatera strateška vprašanja (Wortmann & Flüchter, 2015). Da pa bo vse našteto imelo pravilni in pričakovani učinek, bo v prihodnosti naraščala tudi potreba po kadru, veščem na področju analiziranja masovnih podatkov (Chaudhari & Patel, 2017).

Na žalost pa treba poudariti tudi, da bo v prihodnosti zasebnost podatkov postala še večji izziv in težava (Marr, 2016a). Pri Gartnerju predvidevajo, da se bo do leta 2018 50 odstotkov kršitev poslovne etike zgodilo skozi neprimerno uporabo masovnih podatkov (Goasduff, 2015).

2 VIZUALIZACIJA MASOVNIH PODATKOV

2.1 Zgodovina vizualizacije podatkov

Vizualizacija podatkov ima dolgo zgodovino, ki sega vse do prvih poslikav jamskega človeka. Tekom stoletij so se načini vizualizacije podatkov precej spreminjali in napredovali.

V 17. stoletju so se ljudje največ ukvarjali z vizualizacijo fizičnih meritev, kot so čas, pot in prostor (Friendly, 2008). Prikaz kvantitativnih podatkov v dvodimenzionalnem prostoru, kar danes imenujemo grafikon, je v 17. stoletju odkril Rene Descartes (Few, 2009). V istem stoletju je Michael Florent van Langren ustvaril verjetno prvo vizualno prezentacijo statističnih podatkov (Friendly, Valero-Mora, & Ulargui 2010). Langren je vizualiziral določitve razdalje od Toleda do Rima, čeprav bi to lahko predstavil v tabelarni obliki. Vendar lahko le grafična predstavitev prikaže veliko razlikovanje v ocenah (Friendly, 2008).

V 18. stoletju so predvsem na področju kartografije ustvarjalci poizkušali prikazati več kot samo geografski položaj na zemljevidu. Rezultat tega je, da so se razvili novi načini vizualizacij tudi z vključevanjem fizičnih podatkov. Proti koncu 18. stoletja opazimo tudi prve poizkuse vizualiziranja geoloških, gospodarskih in zdravstvenih podatkov (Friendly, 2008).

Friendly (2008) nadalje povzema zgodovino proti koncu 18. oziroma na začetku 19. stoletja, ko je William Playfair razvil večino grafikonov, ki se uporabljajo še danes, kot so na primer stolpični grafikon in strukturni krog. Z razvojem omenjenih vizualizacijskih tehnik se je v prvi polovici 19. stoletja pojavila eksplozija statističnih grafikonov in tematskih vizualizacij. V tem času so tudi v strokovnih krogih prepoznali koristnost grafičnih prikazov za gospodarsko in strateško načrtovanje.

Few (2009) navaja, da je na razvoj metod in novih načinov vizualizacij proti koncu 20. stoletja močno vplivalo podjetje Apple Computer, ki je predstavilo prvi priljubljeni in ljudem cenovno dosegljiv računalnik, ki se je osredotočal na grafiko, kot orodje za interakcijo in prikaz. To je bil tako rekoč mejnik za uporabo vizualiziranih podatkov in interakcijo z njimi s pomočjo računalnika.

Na prehodu iz 20. stoletja v 21. se je računalniški in programski razvoj, vezan na vizualizacijo podatkov, še krepil in naredil informacijsko tehnologijo dostopnejšo čedalje večjemu krogu ljudi, kar je vplivalo tudi na vizualizacijske tehnike (Few, 2009). Med njimi so to bile nove metode za prikaz večdimenzionalnih podatkov, razvoj interaktivnih orodij, čedalje večja pozornost pa se je začela namenjati tudi dojemanju in zaznavanju vizualiziranih podatkov (Friendly, 2008).

Najnovejše metode vizualizacij se osredotočajo na prikaze ogromnih količin podatkov v specifičnih specializiranih sobah, v katerih so stene opremljene z velikimi zasloni in projektorji. To omogoča znanstvenikom, da analizirajo visoko ločljive dvo- in tridimenzionalne podatke v virtualni resničnosti (Febretti et al., 2013).

Few (2009) navaja, da so dandanes vizualizacije podatkov pomemben del poslovnih inteligenčnih sistemov. Vizualizacije podatkov so dejansko začeli uporabljati in s pridom izkoriščati za poslovne ter vsakodnevne aktivnosti. Opažen je tudi trend, da bodo največje

koristi vizualiziranih podatkov prišle skozi analitiko. Vizualna analitika (angl. *visual analytics*) ne prikaže samo podatkov, temveč omogoča, da lahko z njimi tudi interaktivno upravljamo.

2.2 Vizualizacija masovnih podatkov

Namen vizualizacije podatkov je pridobiti vpogled v različne vidike, ki nas zanimajo, s pomočjo različnih interaktivnih orodij (Telea, 2014). Prav tako je eden izmed glavnih namenov sporočati informacije jasneje in učinkoviteje s pomočjo grafičnih orodij (Dur, 2012).

Vizualizacija masovnih podatkov ima marsikatere prednosti. Nekatere izmed njih so (Olshannikova et al., 2015; Ware, 2013):

- razkritje vzorcev znotraj podatkov;
- sprotna interakcija z vizualizacijo;
- zmožnost zaobjeti veliko količino podatkov, iz katerih se lahko hitro razbere pomembnejše od manj pomembnih;
- spodbuda za razumevanje značilnosti tako velikih kot manjših količin podatkov;
- priložnost, da izzivi, povezani s podatki, postanejo takoj vidni;
- informacija tudi o tem, kako so podatki bili zbrani.

Uporaba naprednejših tehnik vizualizacije podatkov ima med drugim veliko prednosti tudi v poslovnem svetu. Podjetja lahko hitro dobijo vpogled v vse stranke oziroma uporabnike in tako izboljšajo svojo storitev oziroma razmerje s strankami. Natančneje lahko segmentirajo stranke s pomočjo podatkov, ki jih dobijo iz zgodovine nakupov. Vizualizacija podatkov omogoči tudi spremljanje finančnih trendov poslovanja, uporablja pa se lahko tudi za analizo tveganj (Olshannikova et al., 2015). Vizualizacija kot orodje nam tako omogoča boljše razumevanje podatkov. Omenja se celo kot naslednja značilnost oziroma naslednji veliki v masovnih podatkov (Kung, 2017).

Vizualizacijo podatkov lahko vidimo tudi protislovno: na eni strani nam ponuja širok in takojšnji vpogled v kopico znanja, ki nam ga prinašajo masovni podatki, na drugi strani pa z vizualizacijo podatkov preizkušamo sposobnost vizualizacije masovnih podatkov (McCosker & Wilken, 2014). Poudariti je treba, da bodo ljudje morali podatke čedalje bolj spreminjati v uporabno obliko, da bi jih v celoti razumeli (Berry, 2011).

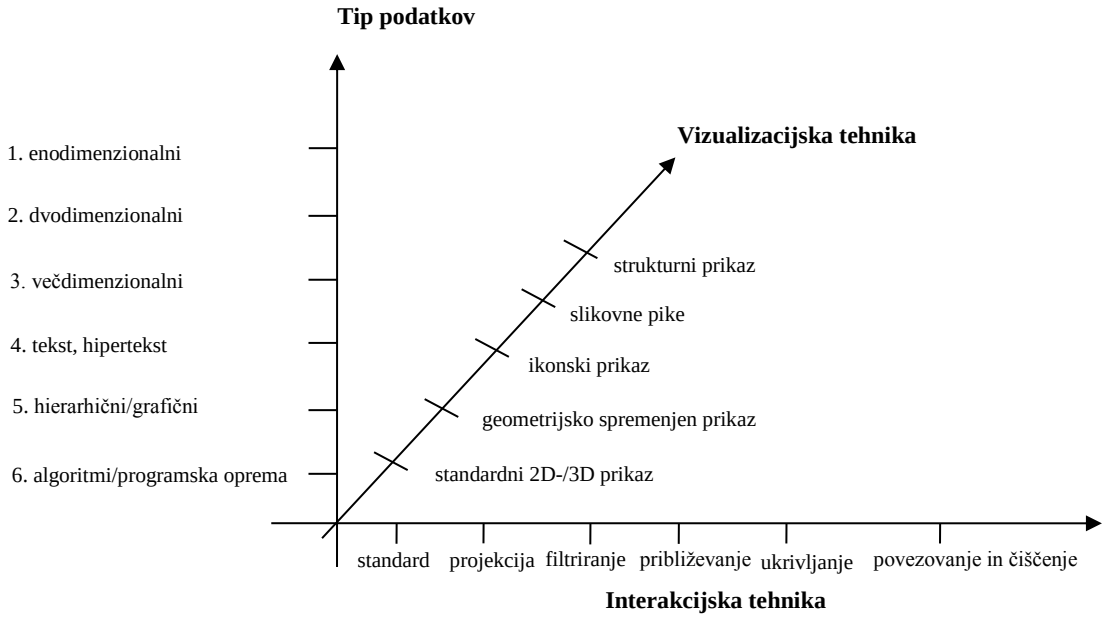
Podajanje sporočilnosti prek vizualizacije podatkov ni lahko. Ni namreč samoumevno, da vizualizacija podatkov reši vse zagate razumevanja teh podatkov. Zato je treba kritično razmišljati, kako vizualizacije pravzaprav sploh delujejo (Kennedy & Hill, 2016).

2.2.1 Vizualizacijske tehnike

Obstaja več različnih delitev vizualizacijskih tehnik. Tehnike so lahko klasificirane glede na tri različne kriterije, kot je prikazano tudi na Sliki 2 (Keim, Hao, Dayal & Hsu, 2002; Olshannikova, et al., 2015):

- glede na tip podatkov;
- glede na vizualizacijsko tehniko;
- glede na uporabljeno interakcijsko tehniko.

Slika 2: Klasifikacija vizualizacijskih tehnik



Vir: D. Keim et al., Pixel bar charts: a visualization technique for very large multi-attribute data sets, 2002, str. 21.

Prvi kriterij klasifikacije vizualizacijskih tehnik je torej tip podatkov, ki jih bomo vizualizirali. Ali so to enodimenzionalni podatki, kot so časovne vrste, dvodimenzionalni, kot so na primer zemljepisne koordinate ali zemljevidi, večdimenzionalni, kot so na primer finančni indikatorji, tekst in hipertekst kot na primer časopisni članki, hierarhični in grafični podatki kot na primer telefonski pogovori, algoritmi in programska oprema (Olshannikova et al., 2015; Keim et al., 2002).

Drugi kriterij klasifikacije so vizualizacijske tehnike. Tu ločimo standardne dvo- ali trodimenzionalne prikaze, kot so na primer linijski in stolpčni grafikoni, katerih glavna slabost je težaven prikaz kompleksnih podatkovnih vrst (Olshannikova et al., 2015; Keim, et al., 2002). Geometrijske transformacije (angl. *geometric transformations*) prikažejo podatke v razsevnem diagramu (Olshannikova et al., 2015). Ikonski prikaz (angl. *icon-based*

displays) vizualizira vsak del podatkov v obliki ikon, podatkovne vrednosti pa kot poteze ikon (Keim et al., 2002).

Za tehniko slikovnih pik (angl. *dense pixel display*) in strukturni prikaz (angl. *stacked display*) Olshanikova et al. (2015) ugotavljajo, da tehnika slikovnih pik prikaže podatke v barvastih slikovnih pikah, ki jih nato smiselno združi skupaj, strukturni prikaz pa se uporabljajo za hierarhično strukturirane podatke in podatke vizualizira v obliki drevesnih struktur.

Nekateri strokovnjaki so vizualizacijske tehnike razvrstili glede na dimenzionalne vidike v vizualizaciji podatkov in tako tehnike razdelili v sedem do osem osnovnih skupin. In sicer: enodimenzionalne, dvodimenzionalne, tridimenzionalne, večdimenzionalne, drevesne, mrežne in na delovni prostor (Eden, 2005).

Vizualizacijske tehnike lahko razdelimo tudi glede na strategije, potrebne za prikaz. Te so (Heo & Hirtle, 2001; Žumer & Merčun, 2008):

- **tehnika ukrivljenja** (angl. *Distortion Technique*), kamor lahko vključimo tehniko ribjega očesa (angl. *Fisheye Eye),* ki želi poudariti en ali več pogledov. Poglede, ki so v središču zanimanja, torej poveča proporcionalno z njihovo pomembnostjo, ostala področja pa so zmanjšana;
- **tehnika približevanja** (angl. *Zoom Technique*) nam dele, ki nas zanimajo, enakomerno poveča, preostalo pa odstrani iz pogleda;
- **tehnika razširjenja pogleda** (angl. *Expanding Outline Technique*) nam prikaže različne ravni podatkov z uporabo razširjene drevesne strukture (angl. *expandable tree*), pri čemer je najpomembnejši del prikazan pri koreninah;
- **tehnika tridimenzionalnega prikaza** (angl. *Three-Dimensional Layout Technique*) želi uporabniku omogočiti vizualno izkušnjo v resničnem svetu.

2.2.2 Vizualizacijska orodja

Pri masovnih podatkih je še posebno težko izvesti vizualizacijo podatkov zaradi njihove velikosti in dimenzije. Trenutna vizualizacijska orodja masovnih podatkov so večinoma slabo zmogljivostna, s slabimi funkcijami in zmožnostmi prilagajanja ter sprotne odzivnosti (Chen & Zhang, 2014).

Ustvarjalci vizualizacij morajo dobro premisliti, katero orodje bodo uporabili (Kennedy & Hill, 2016). Dandanes je na voljo veliko javno dostopnih orodij in aplikacij, ki lahko ustvarijo smiselne in privlačne vizualizacije (Olshannikova et al., 2015). Nekatere izmed njih, ocenjene tudi kot najboljša vizualizacijska orodja v letu 2016, so: Microsoft Power BI, Tableau Desktop, Looker, Domo, Qlik Sense Enterprise Server, Zoho Reports, Information

Builders WebFocus, Tibco Spotfire in Charti (Rist, 2016). Med drugimi pa Warden (2011) omenja še ostale, kot so Gephi, GraphViz, Processing, Protovis in Fusion Tables.

2.3 Vodila in priporočila za dobro vizualizacijo podatkov

2.3.1 Načela dobre vizualizacije

Vizualizacije so pretvornik med kompleksnim svetom podatkov in človekom (Wiener, 1964). Z njihovo pomočjo lahko končnemu uporabniku pomagajo raziskovati, opisovati ali pojasnjevati obravnavan pojav (Kirk, 2016). Najboljše vizualizacije nam pomagajo odgovoriti na pomembna vprašanja in podpirati pomembne odločitve (Wills, 2012).

Načela dobre vizualizacije lahko opredelimo kot (Kirk, 2016; Kulyk, Kosara, Urquiza, & Wassink, 2007; Wills, 2012; Zhu, 2007):

- preprostost;
- interaktivnost;
- prefinjenost;
- točnost;
- učinkovitost;
- uporabnost;
- zanesljivost.

Vizualizacija mora biti preprosta za uporabo in prepoznavanje. Končni uporabnik ne sme porabiti preveč časa samo za ugotavljanje, kako vizualizacija deluje (Kulyk et al., 2007). Ni nujno, da je dobra vizualizacija vedno inovativna. Seveda je treba, da se za reševanje določenih izzivov včasih uporabi tudi inovativnost (Kirk, 2016). Prevelika inovativnost pri vizualizacijah namreč lahko zmede končnega uporabnika in ga odvrne od glavne podane teme prikaza.

Interaktivnost je pomemben element sodobnih vizualizacij. Interaktivne vizualizacije omogočajo uporabniku, da lahko spreminja vhodne podatke oziroma parametre, prikazane na vizualizaciji, pri čimer se te spremembe takoj pokažejo in spremenijo na sami vizualizaciji (Wills, 2012). Pri interaktivnih vizualizacijah lahko izločimo za nas nepomembne dele in vidimo podrobnosti na tistih delih vizualizacije, ki nas še posebej zanimajo (Few, 2009).

Nekatere vizualizacije vključujejo preveč, druge premalo detajlov (Kulyk et al. 2007). Paziti moramo, da končnega uporabnika ne obsujemo s preveč elementi, za katere bi se na koncu izkazalo, da so nepotrebni in da celo kvarijo končno sliko. Da se izognemo vizualizaciji nepotrebnih podatkov, je Tufte (1983) predstavil preprost kriterij oziroma razmerje »podatki

– črnilo« (angl. *data – ink ratio*). Glavni cilj je maksimirati to razmerje, tako da odstranimo nebistvene elemente iz vizualizacije.

Učinkovitost vizualizacije je naslednji pomemben element, kar pomeni, da moramo pri ustvarjanju vizualizacije stremeti k temu, da končni uporabnik iz vizualizacije lahko izlušči podatke, ki jih išče (Kulyk et al., 2007). Bistvo vizualiziranja podatkov ni v tem, da stvari poskušamo olepšati in narediti prezentacijo lepo, da bi očarali občinstvo (Few, 2009). Na koncu je najpomembnejše, če končni uporabnik z lahkoto razume pomen vizualizacije masovnih podatkov, ne da bi poznal celotno analizo podatkov v ozadju (Xu & Shi, 2015). Pri vizualizacijah želimo preprosto pokazati, kaj je pomembno (Wills, 2012).

Ne smemo pa pozabiti niti na izbiro najprimernejše vizualizacijske tehnike za uspešno vizualizacijo masovnih podatkov (Chen & Zhang, 2014). Za učinkovito vizualizacijo je običajno treba uporabiti tudi interakcijske tehnike. Te omogočajo končnemu uporabniku, da je v neposrednem stiku z vizualizacijo (Keim et al., 2002).

Za transparentnost s končnim uporabnikom vizualizacije Kirk (2016) ugotavlja, da je ta ključna, zato je dobro razmisliti, da z njim delimo informacijo o tem, kako so bili podatki zbrani, ali smo v procesu vizualizacije kaj izključili in kakšne kalkulacije ali spremembe smo uporabili. Pri ustvarjanju vizualizacije nam mora biti glavno vodilo to, da nikoli ne zavajamo končnega uporabnika vizualizacij. Moramo se izogibati vsaki možnosti napačne interpretacije, nesporazuma in netočnosti prikazanih podatkov.

Pri ustvarjanju vizualizacije je torej dobro že na začetku premisliti o naslednjih dejavnikih (Wills, 2012):

- o **predmetu prikaza**, pri čimer je treba pred vizualizacijo dobro proučiti, kar bomo pozneje prikazovali;
- o **namenu vizualizacije**, pri čemer moramo vnaprej pomisliti, zakaj je vizualizacija sploh potrebna in kakšne ukrepe pričakujemo od uporabnikov te vizualizacije;
- o tem, **kdo je uporabnik vizualizacije**, pri čimer je izjemno pomembno, da ne samo, da poznamo končnega uporabnik vizualizacije, ampak tudi, kako in za kaj bo uporabljena.

2.3.2 Človekovo dojemanje vizualizacije

Poznavanje človekovega zaznavanja in dojemanja vizualizacije je pomembna komponenta dobre vizualizacije (Görg, Pohl, Qeli, & Xu, 2007), pri čemer moramo imeti v mislih tudi omejenost človekove percepcije (Kulyk et al., 2007). Pri človeškem dojemanju vizualizacij govorimo o treh fazah, ki so (Kirk, 2016):

- zaznavanje (angl. *perception*): kaj vizualizacija prikazuje;

- interpretacija (angl. *interpretation*): kaj prikazano pomeni;
- razumevanje (angl. *comprehension*) : kaj prikazano pomeni meni.

Kirk (2016) ugotavlja, da zaznavanje vizualizacije pomeni preprosto zmožnost branja vizualizacije in vprašanje, kaj vizualizacija prikazuje. Interpretacija je naslednji korak dojemanja vizualizacije, ko uporabnik vizualizacije poizkuša iz videnega sestaviti nekaj smiselnega. Nazadnje se uporabnik sprašuje še o tem, kaj lahko naredi s prikazanim in ali je vizualizacija dala nek nov vpogled v prikazano temo. Uporabnik se prav tako sprašuje, ali iz vizualizacije izhajajo določene prihodnje naloge ali ukrepi, po katerih mora poseči.

Za zaznavanje in razumevanje vizualizacije so pomembne tudi vizualne spremenljivke, kot so položaj, oblika, usmerjenost, barva, struktura, vrednosti, velikost ter gibanje (Bertin, 1983; Görg, et al., 2007). Izbira vizualnih spremenljivk namreč močno vpliva na to, kako bodo uporabniki vizualizacije dojemali in uporabili prikazane podatke (Görg, et al., 2007).

Kirk (2016) meni, da je za dobro interpretacijo vizualizacije ključni dejavnik uporabnikovo poznavanje prikazanega objekta. V primerih, da uporabnik nima predhodnega znanja o prikazanem objektu, morajo ustvarjalci vizualizacije z jasno vizualizacijo pomagati, da bo uporabnik lažje razbral pomen prikazanega. Na to, kako bodo različni končni uporabniki interpretirali vizualizacijo, vpliva veliko dejavnokov. Nekateri izmed pomembnejših so raven interesa do prikazane tematike, njena relevantnost za uporabnika in pripravljenost uporabnika za interpretacijo v danem trenutku.

Eno pomembnejših vlog pri dojemanju vizualizacije ima pravilna izbira barve, pri čemer morajo imeti ustvarjalci vizualizacij nekaj razumevanja o njihovem učinku. Barve se pogosto uporabljajo zato, da ustvarijo oziroma poudarijo nek pomen, ne pa za dekoracijo (Kennedy & Hill, 2016). Z barvami lahko dodamo pravšnji pomen določenemu predmetu, ki ga prej ni imel. Že v vsakdanjem življenju imajo barve pomembno vlogo na način, da ali poizkušajo vzbuditi pozornost ali pa ravno obratno, nas želijo odvrniti od neke stvari (Andersen, Vuori, & Guillaume, 2015).

2.3.3 Pripovedovanje zgodbe skozi vizualizacijo

Zgodbe so zelo prikladne, saj jih lahko uporabimo za prenos informacij, izkušenj, idej in svojih vpogledov (Wojtkowski & Wojtkowski, 2002).

Dandanes lahko opazimo, da praktično vsi, od medijev do politikov, uporabljajo vizualizacijo kot temelj in za poudarek zgodbe, ki jo želijo podati (Segel & Heer, 2010). Uporaba modernih vizualizacijskih orodij pomaga pri podajanju zgodbe nestrokovnjakom obravnavanega področja (Lundblad & Jern, 2012). Nekatera orodja za vizualizacijo podatkov so namreč že začela vključevati elemente pripovedovanja zgodbe v svoje aplikacije (Segel & Heer, 2010).

Vizualizacija je lahko močno sredstvo, vendar da hkrati pove še neko zgodbo, mora biti opremljena še z ostalimi komponentami, kot so besedilo, zvok, slika, posnetek (Kosara & Mackinlay, 2013). Vizualizacija je lahko tudi zelo interaktivna, kar prenaša breme pripovedovanja zgodbe skozi vizualizacijo tako na avtorja kot uporabnika te vizualizacije (Segel & Heer, 2010). Pripovedovanje zgodbe skozi vizualizacijo je torej ključnega pomena za podajanje intuitivnih in hitrih vpogledov v ogromne količine podatkov z njihovo sprotno obdelavo (Wojtkowski & Wojtkowski, 2002).

Jeffery in Al-Gharaibeh (2015) ugotavljata, da je, če vizualizacija ne uspe podati želene zgodbe, to skoraj huje kot neobstoj vizualizacije na prvem mestu. Precej hitro se lahko zgodi tudi izguba kritičnih informacij v vizualizaciji. Pri tem uporabnik vizualizacije vedno niti ne ve, da kritični del informacij ni prikazan. Situacija pa ni nič boljša, če uporabnik vizualizacije ne zna razbrati.

2.4 Prihodnost vizualizacij masovnih podatkov

Opazimo lahko trend, da se vizualizacije podatkov razvijajo skupaj z napredkom v tehnologiji. Posledica tega so in bodo premiki vizualizacij podatkov proti čedalje več vključenim animacijam in interaktivnim vizualizacijam (Brenner, 2016). Končni uporabniki bodo namreč čedalje več želeli odkrivati sami (Brown, 2015).

V prihodnosti bo treba vizualizacije čedalje bolj prilagajati specifičnim strankam glede na njihove potrebe (Becker, 2016). Umetniki, oblikovalci in psihologi se čedalje bolj posvečajo raziskovanju možnosti in idej, da bi ustvarjali vizualizacije, ki bi pripovedovale neko zgodbo in delile izkušnjo (Segel & Heer, 2010). Področje vizualizacij se premika od umetnosti čedalje bolj proti znanosti (Towler, 2015).

Pomembna komponenta v prihodnosti vizualizacije podatkov so tudi vizualizacije v virtualni resničnosti. Virtualna resničnost predstavlja revolucijo v vizualizaciji predvsem masovnih podatkov (Krystian, 2017). S postavitvijo človeka v digitalno ustvarjen prostor, z možnostjo 360-stopinjskega vpogleda in s simulacijo gibanja v treh dimenzijah bi lahko še bolje izkoristili širino podatkov, ki jih nosijo naši možgani (Marr, 2016b).

Za prihodnost obstaja tudi veliko nadaljnjih izboljšav na področju vizualizacije masovnih podatkov. Kljub temu da se je področje že precej razvilo, pa obstaja še veliko izzivov oziroma možnosti izboljšav za prihodnost (Comai, 2015):

- vizualizacije morajo biti vizualno privlačne, prepričljive in močno orodje za komunikacijo;
- vizualizacije morajo biti povezane in večdimenzionalne;
- uporaba algoritmov in umetne inteligence za simuliranje prihodnjih dogodkov;

- vizualizacijska orodja morajo biti zasnovana tudi za mobilne telefone, tako da lahko vizualizacijo ustvarimo v vsakem okolju, takoj ko se pojavijo podatki.

Svet postaja čedalje bolj povezan in soodvisen, zato bodo možnosti za ustvarjanje dodane vrednosti skozi vizualizacijo podatkov samo še naraščale. Vizualizacija podatkov bo imela čedalje bolj pomembno vlogo v podjetjih in v družbi (Towler, 2015).

2.5 Vizualizacija poti nekoč in danes

Ker v empiričnem delu svojega magistrskega dela uporabim tri dejanske vizualizacije masovnih podatkov, ki bodo na nek način prikazovale pot oziroma prehajanja, je smotrno, da na kratko predstavim vizualizacijo poti nekoč in danes.

Če se ozrem nazaj v zgodovino, lahko najdem nekaj zelo dobrih primerov vizualizacij, ki so med drugim bogato prikazovale tudi komponento potovanja oziroma prehajanja. Klasičen in eden bolj znanih primerov vizualizacije poti oziroma premikov je vizualizacija Charlesa Josepha Minarde, ki je vizualiziral potovanje Napoleonove vojske in njene izgube v letu 1812 (Chen, 2006). Poleg glavnega sporočila o izgubah Napoleonove vojske Wills (2012) ugotavlja, da vizualizacija bogato prikaže tudi marsikatero drugo komponento, saj je zgrajena na več podatkovnih naborih, ki prikazujejo geografski relief, potovanje vojske, velikost vojske, temperaturo zraka in časovno komponento. Potovanje, vizualizirano v obliki črte, je na tej vizualizaciji označeno in ločeno tudi z različnimi barvami, ki ločijo med tem, ali je vojska potovala naprej ali se je umikala.

V strokovni literaturi se lahko med drugim najde tudi zelo bogata, vendar precej kompleksna vizualizacija poti oziroma vlakov na progi v Indoneziji. Vizualizacija je nastala v prvi polovici 20. stoletja in prikazuje potovanje vlakov po železniški progi v Indiji v 24 urah. Bogato prikazuje potovanje vlaka skupaj z drugimi spremenljivkami, kot sta čas in prostor (Tufte, 1990).

Na kratko lahko povzamem, da se je v zgodovini za prikazovanje poti oziroma premikov največ uporabljal grafični dvodimenzionalni prikaz s povezovanjem različnih točk. Žal pa takšni grafični prikazi ne zadoščajo več popolnoma čedalje večjim količinam podatkov in kompleksnih mrežnih povezav (Partl et al., 2016).

Grafični prikazi so tudi danes pogosta tehnika za prikazovanje povezav med ločenima enotama (Sheny & Maz, 2007). Običajno se pot oziroma povezave prikazujejo s pomočjo povezave vozlišč (angl. *Node-link diagram*), ker je na takšen način relativno lahko prepoznati poti in povezave, saj so te prikazane kot zaporedna povezava linijskih segmentov (Partl et al., 2016; Sheny & Maz, 2007).

Ne smemo pa pozabiti, da kompleksne mrežne povezave vključujejo tudi bogato vrsto podatkov in atributov, ki so pridruženi prikazanim vozliščem in njihovim robovom. Takšni atributi so pogosto kritični za razumevanje prikazane poti, pa vendar njihovo neposredno prikazovanje v vizualizaciji prinese na plan težavo razširljivosti (angl. *scalability*). Ena izmed vizualizacijskih tehnik, ki se sooči s tem izzivom, je Pathfinder, ki nam omogoča prikazovanje poti, izhajajočih iz velikih in kompleksnih mrežnih povezav (Partl et al., 2016).

3 EMPIRIČNA RAZISKAVA O RAZUMLJIVOSTI VIZUALIZACIJ MASOVNIH PODATKOV

3.1 Namen in cilji raziskave

Glede na zapisano v teoretičnem delu magistrskega dela bi nam vizualizacija morala olajšati vpogled v masovne podatke. Na lažji način bi nam morala prikazati vzorce in trende sicer kompleksnih podatkovij. Ali je res tako in kako blizu so ljudem izbrane vizualizacije masovnih podatkov, raziskujem v empiričnem delu magistrskega dela.

Namen raziskovalnega dela je na majhnem vzorcu raziskati, v kolikšni meri so vizualizacije masovnih podatkov razumljive in kako jasno podajajo glavno temo prikaza, torej ali pripomorejo k boljšemu razumevanju sporočila vizualizacije ali vodijo do napačnega razumevanja. Predstaviti želim tudi napotke za nadaljnje raziskovanje, na primer kako vizualizacije masovnih podatkov z vidika podajanja glavne teme prikaza narediti še razumljivejše.

Dva glavna cilja empiričnega dela magistrskega dela na osnovi kvalitativne analize pridobljenih primarnih podatkov sta:

- ugotoviti, do kolikšne mere so izbrane vizualizacije masovnih podatkov razumljive;
- ugotoviti, kako jasno vizualizacije masovnih podatkov podajajo svojo glavno temo prikaza.

V empiričnem delu magistrskega dela hočem torej proučiti, do kolikšne mere so vizualizacije masovnih podatkov razumljive in kako jasno podajajo svojo glavno temo prikaza.

3.2 Raziskovalne hipoteze

Za proučevanje razumljivosti in jasnosti vizualizacij je najbolj primerna kvalitativna obravnava, ki uporabniku omogoča vizualizacije, da izrazi svoje poglede, prepričanja in tudi dvome. Glede na zastavljena glavna cilja postavljam več raziskovalnih hipotez, ki pa niso namenjene statističnemu preverjanju, ampak jih gre razumeti v luči sistematičnega iskanja

kvalitativnih podatkov, ki bodisi podpirajo bodisi zavračajo posamezno hipotezo. Uporaba izraza preverjanje hipotez je zato mišljena kot vrednotenje teh hipotez.

- **Hipoteza 1:** Vizualizacijo masovnih podatkov gledamo skozi tradicionalne metode grafičnih prezentacij.
- **Hipoteza 2**: Vizualizacije so hitreje razumljive mlajšim udeležencem, saj so bolj izpostavljeni novejšim oblikam vizualizacije podatkov.
- **Hipoteza 3:** Predznanje z določenega področja izboljša razumevanje vizualizacij masovnih podatkov s tega področja.
- **Hipoteza 4:** Pomanjkljiva vizualizacija masovnih podatkov lahko poda napačno sporočilnost.
- **Hipoteza 5:** Če nekdo najprej sliši zgodbo in šele nato vidi s to zgodbo povezano vizualizacijo, bo hitreje dojel in bolje razumel vizualizacijo, kot če vidi samo vizualizacijo in poizkuša ugotoviti bistvo prikaza.

3.3 Metodologija raziskave

Empirični del magistrskega dela temelji na kvalitativnem raziskovanju, ki bo v manjši meri tudi kvantitativno podprto. S pomočjo treh specifičnih vizualizacij masovnih podatkov, ki se med sabo razlikujejo glede na težavnost in način predstavitve, iščem argumente za začetno postavljene hipoteze in proti njim. Udeleženci bodo s pomočjo različnih nalog ugotavljali, v kolikšni meri analizirani primeri vizualizacij masovnih podatkov prikazujejo glavno temo in specifična sporočila.

Izvedem kombinacijo delno strukturiranega intervjuja in opazovalne študije, prevladuje slednja. Kombinacija teh dveh metod zbiranja primarnih podatkov omogoča pridobiti precej podrobne odgovore. Zbiranje kvalitativnih podatkov mi omogoča pridobivanje podrobnih podatkov za analizo, kljub temu da je ta metoda nekoliko zamudnejša in je lahko podvržena določeni stopnji subjektivnosti (Bregar, Pfajfar, & Ograjenšek, 2009).

Z delno strukturiranim intervjujem bom dobila odgovore na vnaprej pripravljena konkretna vprašanja v zvezi z raziskovalnim problemom. S pomočjo opazovalne študije pa bom lahko opazovala obnašanje udeležencev; predvsem bom lahko spremljala, kako se bodo lotili vizualizacij.

Uporaba opazovalne študije kot način pridobivanja primarnih podatkov je zelo primerna za pridobitev celovite in poglobljene slike pojava, vedenja in podobno, metoda pa je običajno cenejša od ostalih metod zbiranja primarnih podatkov (Bregar, Ograjenšek, & Bavdaž, 2005). Po drugi strani pa izvedba opazovanja zahteva precej časa, težje je tudi zagotoviti objektivnost (Mack, Woodsong, MacQueen, Guest, & Namey, 2005). Uporaba opazovalne študije v praksi je povezana tudi z etičnimi dilemami, kot je na primer zagotovitev pravice opazovane osebe do zasebnosti pri opazovanju (Bregar et al., 2009).

V primeru tega magistrskega dela zgoraj naštete prednosti opazovalne študije prav gotovo veljajo. Glede pomanjkljivosti, ki se dotika etičnih dilem, pa bom pri empiričnem delu udeležence že vnaprej seznanila s tem, da bodo njihovi odgovori zaupni, nikjer individualno razkriti in uporabljeni samo za namene mojega magistrskega dela.

Z opazovanjem bom dobila celovit in poglobljen vpogled v to, kakšni bodo ekspresivni odzivi udeležencev pri reševanju nalog, opazovala bom tudi, koliko časa bodo udeleženci porabili za posamezno nalogo. Z opazovalno študijo bom dobila celovito sliko o tem, kaj udeležencem povzroča največ težav. Uporabila bom strukturiran način opazovanja, saj si bom opažanja zapisovala v vnaprej pripravljeno tabelo, sama pa bom neprikriti opazovalec. Pogovor z udeleženci bom – z njihovo privolitvijo – tudi snemala z mobilnim telefonom. Pridobljene podatke bom na ta način pozneje lažje analizirala.

Pri intervjuju govorimo o pogovoru z določenim namenom (Bregar at al., 2005). Prednost intervjuja na splošno je, da je neizogibno interaktiven. Spraševalec lahko že med intervjujem postavi dodatna vprašanja ali preveri, ali je izpraševanca pravilno razumel. Prav tako lahko spraševalec že med intervjujem spremlja odziv izpraševanca in njegovo neverbalno komunikacijo (Miller & Glassner, 2011; Vogrinc, 2008).

Ena izmed slabosti intervjuja je, da lahko odnos med spraševalcem in izpraševancem vpliva na odgovore slednjega. Ko se udeleženca intervjuja poznata, lahko to vpliva na kakovost in iskrenost odgovorov izpraševanca. Na drugi strani pa lahko tudi sam spraševalec zavedno ali pa nezavedno vpliva na odgovore izpraševanca (Miller & Glassner, 2011; Vogrinc, 2008).

V empiričnem delu magistrskega dela bom lahko dobro izkoristila prednost, ki jo ponuja intervju, saj bom lahko spremljala neverbalni odziv izpraševancev, prav tako pa tudi interaktivnost, ki jo intervju prinaša. Po drugi strani bom poizkušala izničiti svoj negativni vpliv na odgovore izpraševanca tako, da ne bom z neverbalno komunikacijo izražala lastnega mnenja o odgovorih izpraševancev, ampak se bom čimbolj nevtralno odzivala.

Intervju mi bo prinesel odgovore na to, kako udeleženci dojemajo vizualizacije masovnih podatkov, kaj se jim bo zdelo težje in kaj lažje pri določenih nalogah ter specifičnih vizualizacijah masovnih podatkov.

Postavljenih raziskovalnih hipotez ne bom preverjala s statističnimi metodami in orodji, saj bodo pridobljeni podatki v veliki meri kvalitativni. Analizo kvalitativnih podatkov bom izvedla z analitičnimi orodji, kot so zapiski in prepisi posnetkov, ki bodo nastali med empirično raziskavo, pa tudi povzetki, ki bodo pripravljeni na osnovi zapiskov in prepisov.

Pri urejanju pridobljenih kvalitativnih podatkov bom sledila procesu njihovega urejanja. Proces sestavljajo naslednji koraki: seznanjenje, vrednotenje, konceptualizacija, kategorizacija konceptov, ponovno kodiranje, povezovanje in ponovno vrednotenje (Bregar

et al., 2005). Kjer bo to smiselno, bom uporabila numerične podatke in iz njih izračunane statistike.

Pri vrednotenju hipotez si bom torej največ pomagala z vnaprej pripravljenimi tabelami, kamor bom vključila odgovore udeležencev med raziskavo. Pri nekaterih vprašanjih pa bom imela tudi vnaprej pripravljene kriterije odgovorov. Vsebino dejanskih odgovorov bom nato primerjala z začetno postavljenimi kriteriji.

3.4 Potek empirične raziskave

3.4.1 Oblikovanje vprašalnika

Izvedba kombinacije delno strukturiranega intervjuja in opazovalne študije bo temeljila na treh dejanskih vizualizacijah masovnih podatkov, ki se bodo med sabo razlikovale glede na težavnost ter način vizualizacije, vseeno pa bodo imele nekaj skupnega. Vsaka bo na svoj način prikazovala poti, tokove oziroma premike.

Ena od vizualizacij bo prikazovala potovanje sporočil na enem izmed družbenih omrežij. Za ta primer sem se odločila, ker se na družbenih omrežjih v vsakem trenutku ustvari ogromna količina podatkov. Naslednja vizualizacija bo prikazovala prehod uporabnikov od enega izdelka k drugemu. Ta vizualizacija izhaja s področja genomike, ki ga prav tako zaznamuje ogromna količina podatkov. Tretja vizualizacija bo prikazovala potovanje izdelkov v proizvodnji, opremljeni z brezžičnimi vmesniki, ki so bogat vir masovnih podatkov. Za ta primer vizualizacije sem se odločila, ker tudi ta prikazuje gibanje oziroma premike, vendar na drugačen način kot ostali dve vizualizaciji. Tako bom lahko ugotavljala, ali vizualizacija pripomore k razumevanju njene sporočilnosti ali ne.

Deset udeležencev, vključenih v raziskavo, je bilo naključno razdeljenih v dve skupini s po pet udeležencev. Za preverjanje začetno postavljenih hipotez empiričnega dela magistrskega dela sem sestavila dva tipa nalog za celoten nabor dejanskih vizualizacij. Posamezna skupina udeležencev ni imela enakega tipa nalog za vsako izmed treh vizualizacij, zato razdelitev udeležencev glede na tip nalog za lažje razumevanje prikazujem v Tabeli 1.

Tabela 1: Razdelitev skupine udeležencev in tipa nalog

Tip naloge	**Vizualizacija 1**	**Vizualizacija 2**	**Vizualizacija 3**
Prvi tip nalog	Skupina 1	Skupina 2	Skupina 1
Drugi tip nalog	Skupina 2	Skupina 1	Skupina 2

Uporabila sem dva tipa nalog, katerih potek na kratko opisujem spodaj, nanju pa se bom navezovala tudi skozi celotno poglavje analize rezultatov empirične raziskave.

- **prvi tip nalog: udeleženci na osnovi opisa sami pripravijo vizualizacijo (vizualizacijska skupina):**
  - udeleženci najprej dobijo opis vsebine, pri čemer dejanske vizualizacije še ne vidijo,
  - na osnovi opisa morajo narediti svojo vizualizacijo,
  - ko končajo s svojo vizualizacijo, jim pokažem še dejansko vizualizacijo,
  - udeleženci morajo nato opisati podobnosti in razlike med svojo in dejansko vizualizacijo ter prednosti in izzive dejanske vizualizacije;
- **drugi tip nalog: na osnovi videne vizualizacije udeleženci ugotavljajo temo prikaza (skupina s ključnimi besedami):**
  - udeleženci najprej dobijo dejansko vizualizacijo, brez razlage, kaj predstavlja, in brez naslova, čeprav se naslov tipično obravnava kot del vizualizacije. Dejansko vizualizacijo jim predstavim brez naslova, ker želim raziskati, kaj udeleženci zmorejo pridobiti samo iz grafičnih elementov,
  - na osnovi predložene dejanske vizualizacije morajo naštevati ključne besede, ki se jih spomnijo ob gledanju vizualizacije,
  - na osnovi ključnih besed morajo udeleženci določiti naslov vizualizacije.

S prvim tipom nalog bom preverjala prvo hipotezo, z drugim tipom nalog bom preverjala hipoteze številka dve, tri in štiri, medtem ko bom peto hipotezo preverjala tako s prvim kot z drugim tipom nalog.

3.4.2 Raziskovalni vzorec

V raziskavo je bilo vključenih deset udeležencev različne stopnje izobrazbe in različne starosti.

Tabela 2: Socialno demografske značilnosti (spol in starost) raziskovalnega vzorca

Spol	**N**
Ženski	6
Moški	4
Skupaj	**10**
Starostni razred	**N**
Od 20 do 25	2
Od 26 do 31	4
Od 32 do 45	2
Od 46 do 51	2
Skupaj	**10**

Tabela 3: Socialno demografske značilnosti (stopnja in smer izobrazbe) raziskovalnega vzorca

Stopnja izobrazbe / smer izobrazbe	**Naravoslovna smer**	**Družboslovna smer**
Srednja poklicna šola	1	0
Visokošolska univerzitetna izobrazba	1	3
Magisterij stroke	0	2
Doktor znanosti	2	1
Skupaj	**4**	**6**

Tabeli 2 in 3 prikazujeta sestavo raziskovalnega vzorca. Starost udeležencev sem oblikovala v frekvenčno porazdelitev. Povprečna starost je 31,7 leta. Stopnja izobrazbe udeležencev je precej visoka: trije imajo doktorski naziv, kar sem vnaprej upoštevala pri sestavi raziskovalnega vprašalnika.

Ker bodo v naslednjem poglavju tudi socialne demografske značilnosti raziskovalnega vzorca imele pomembno vlogo, sem v Tabeli 4 ustvarila kratko legendo, ki jo bom uporabljala v nadaljevanju. V Tabeli 4 sem prilagodila tudi starostne razrede, saj želim v analizi rezultatov empiričnega dela operirati z dvema razredoma oziroma dvema starostnima skupinama. Pričakujem namreč, da imata skupini različne izkušnje na splošno in v povezavi z informacijsko tehnologijo. Pričakujem, da ima mlajša starostna skupina manj življenjskih izkušenj, vendar več izkušenj na področju informacijske tehnologije.

Tabela 4: Kodiranje udeležencev glede na socialno-demografske značilnosti

Udeleženec	**Starostni razred**	**Stopnja in smer izobrazbe**
1	Od 32 do 51	Magisterij stroke (družboslovna smer)
2	Od 20 do 31	Visokošolska univerzitetna izobrazba (družboslovna smer)
3	Od 20 do 31	Doktor znanosti (naravoslovna smer)
4	Od 20 do 31	Visokošolska univerzitetna izobrazba (družboslovna smer)
5	Od 20 do 31	Magisterij stroke (družboslovna smer)
6	Od 32 do 51	Doktor znanosti (družboslovna smer)
7	Od 32 do 51	Srednja poklicna šola (naravoslovna smer)
8	Od 20 do 31	Visokošolska univerzitetna izobrazba (naravoslovna smer)
9	Od 32 do 51	Doktor znanosti (naravoslovna smer)
10	Od 20 do 31	Magisterij stroke (družboslovna smer)

4 ANALIZA REZULTATOV EMPIRIČNE RAZISKAVE

Analiza rezultatov je razdeljena na štiri dele, rezultati treh dejanskih vizualizacij so predstavljeni v podpoglavjih. Ta so nadalje razdeljena na predstavitev rezultatov za oba tipa nalog in povzetek za celotno dejansko vizualizacijo na osnovi pridobljenih rezultatov.

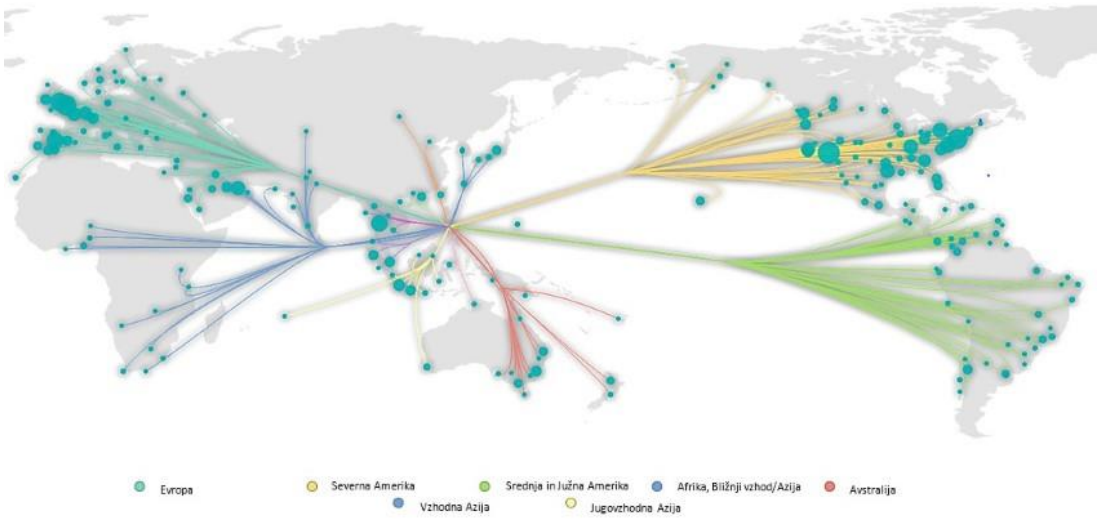
Pred začetkom vsakega podpoglavja se, kot že omenjeno, nahaja dejanska vizualizacija, na kratko jo tudi predstavim. Rezultate primerjam s pričakovanji in domnevami, ki sem si jih postavila pred začetkom vsake naloge pri vsaki vizualizaciji. Pri tem posebej poudarjam, da sem za objavo vizualizacij pridobila dovoljenja avtorjev oziroma založnikov.

V zadnjem delu četrtega poglavja zapišem sklepe in ugotovitve celotne raziskave. Zaključne ugotovitve primerjam z v začetku predstavljenimi petimi hipotezami in jih potem lahko tudi potrdim ali ovržem.

4.1 Analiza rezultatov prve vizualizacije

Vizualizacija na Sliki 4 predstavlja enostaven prikaz pošiljanja tvitov na Filipine iz vseh delov sveta kot izraz pomoči po katastrofalnem tajfunu novembra leta 2013. Vizualizacija prikazuje pot oziroma povezanost svetovnih mest s Filipini, vse se namreč steka v to azijsko državo.

Slika 3: Aktivnost na Twitterju po katastrofalnem tajfunu na Filipinih novembra 2013



Vir: N. Belmonte, Visualised: the world responds to Typhoon Haiyan on Twitter, 2013. Ponatis z dovoljenjem.[1]

[1] Osebna komunikacija z avtorjem Nicolasom Garcio Belmontejem po e-pošti. Dovoljenje po e-pošti pridobljeno z dnem 3. oktober 2016.

4.1.1 Analiza rezultatov vizualizacijske skupine

Udeleženci so na začetku naloge dobili opis dogodka, ki je bil prikazan na prvi vizualizaciji. Opis se nahaja tudi v spodnjem odstavku. Po prebranem so udeleženci morali vizualizirati opisan dogodek, za kar so imeli na voljo tri minute.

»Filipine je novembra 2013 prizadel katastrofalen tajfun. Pomoč in izrazi sožalja ter sočustvovanja so prihajali od vsepovsod, med drugimi je to izrazilo tudi na tisoče ljudi prek Twitterja. To so storili prek tako imenovanih tvitov – kratkih sporočil, kjer so v različnih jezikih iz vseh koncev sveta pošiljali svoje izraze podpore in sočustvovanja. Na listu papirja prikažite pošiljanje teh tvitov na Filipine iz vseh kontinentov.«

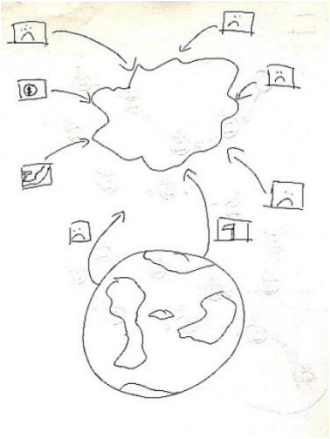
Moja pričakovanja pred začetkom nalog vizualizacijske skupine:

- udeleženci bodo najprej začeli risati zemljevid, nato pa črte oziroma povezave bodisi do Filipinov ali drugače;
- udeleženci bodo poizkušali prikazati simbole Twitterja;
- podobnosti med udeležencevo in dejansko vizualizacijo bodo precejšnje;
- kot prednost dejanske vizualizacije bodo udeleženci navedli, da se zaradi uporabe različnih barv bolje razbere, s katerih lokacij prihajajo tviti;
- kot najmanj razumljive elemente dejanske vizualizacije bodo udeleženci omenili legendo, saj krogci v legendi nimajo nikakršne veze s krogci na dejanski vizualizaciji, prav tako veliko število povezav, ki vodijo na Filipine, in z ničimer ne nakazujejo, kaj pomenijo.

Medtem ko so udeleženci reševali naloge oziroma ustvarjali svojo vizualizacijo, sem merila čas in si zapisovala potek njihovega vizualiziranja. Na koncu sem jih povprašala tudi o podobnostih in razlikah med njihovo in dejansko vizualizacijo.

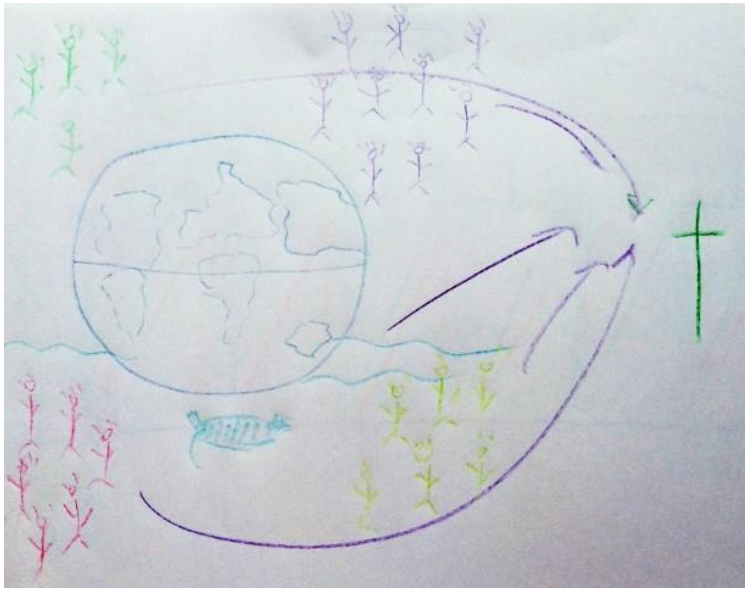
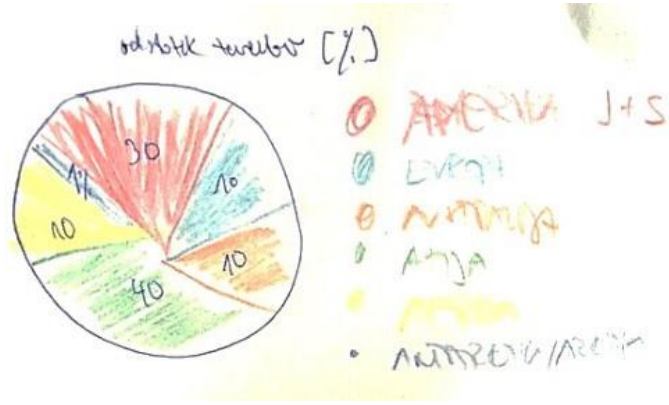
Prav tako so morali našteti, kaj se jim zdi najboljše in najmanj razumljivo pri dejanski vizualizaciji ter ali bi ugotovili tematiko vizualizacije masovnih podatkov, če ne bi poznali predzgodbe.

Tabela 5: Rezultati poteka vizualiziranja prve vizualizacije vizualizacijske skupine

Udeleženec	**Dejanski čas** (mm:ss)	**Potek vizualizacije in končna vizualizacija**
1	02:05	Udeleženec najprej začne risati zemljevid, nato tako imenovane simbole, s katerimi uporabniki označujejo tvite, ki se dotikajo iste teme oziroma pogovora (angl. *hestag*). Označi tudi žarišče dogajanja, Filipine.
2	02:26	Udeleženec najprej nariše Filipine, okrog njih pa s puščico povezane kvadratke, v katerih so bili razni simboli žalovanja. Nazadnje nariše tudi zemljevid sveta, od koder se nato črte stekajo do Filipinov. 
4	02:25	Udeleženec s tipično ikono otoka ponazori Filipine. Doda še mobilni telefon, ki ga poveže s Filipini, in nakazuje, da tako potujejo tviti.

se nadaljuje

Tabela 5: Rezultati poteka vizualiziranja prve vizualizacije vizualizacijske skupine (nad.)

Udeleženec	**Dejanski čas** (mm:ss)	**Potek vizualizacije in končna vizualizacija**
5	03:00	Udeleženec najprej nariše zemljevid sveta. Okrog nariše figure ljudi z različnimi barvami, kar nakazuje, da prihajajo z različnih kontinentov. Nariše tudi ptička, simbol Twitterja. Vse skupinice ljudi poveže v skupno točko, kjer simbol križa nakazuje žalovanje. 
3	02:42	Udeleženec dogodek prikaže v obliki tortnega diagrama. Vsak del diagrama predstavlja kontinent, vsak kontinent je označen z drugo barvo. Vsak košček tortnega diagrama vsebuje tudi odstotek, to je odstotek tvitov, ki so prihajali s posameznih kontinentov. 

Iz Tabele 5 lahko razberem, da so štirje od petih udeležencev najprej začeli risati ali zemljevid sveta ali Filipine. Prav tako so ti štirje udeleženci označili Filipine kot žarišče oziroma osrednji pojem dogajanja, kamor so se nato stekale povezave. Tudi za udeleženca številka tri ne morem trditi, da ni na nek način vizualiziral zemljevida sveta oziroma kontinentalne razdelitve Zemlje. Udeleženci so vizualizacijo torej večinoma začeli s prikazom zemljevida sveta oziroma z vizualizacijo Filipinov. Tako lahko potrdim svoje

pričakovanje, da bodo udeleženci začeli z risanjem zemljevida in nadaljevali s prikazom črt oziroma povezav.

Trije od petih udeležencev so prikazali tudi simbole Twitterja, kot so simboli, s katerimi uporabniki označujejo tvite, ki se dotikajo iste teme oziroma pogovora (angl. *hestag*), modrega ptička ali pa napravo, s katero uporabniki lahko pošiljajo tvite. To nakaže, da se je element prikaza, da vizualizacija govori o pošiljanju tvitov, zdel udeležencem pomemben.

Štirje udeleženci vizualizacijske skupine pri prvi vizualizaciji imajo družboslovno izobrazbo, eden udeleženec naravoslovno. Izmed rezultatov v Tabeli 5 izpostavljam udeleženca številka tri. Ta se je vizualiziranja lotil na drugačen, bolj tehnični način. Ni toliko razmišljal, kako prikazati razne simbole žalovanja, pošiljanja tvitov in podobno, temveč je zgodbo predstavil bolj statistično, v obliki strukturnega kroga in z odstotki.

Udeleženci so kot podobnosti med svojo in dejansko vizualizacijo največkrat navajali:

- prikaz Filipinov kot središče dogajanja;
- ponazoritev, s katerih kontinentov prihajajo tviti;
- zemljevid sveta;
- uporabo barv za razločevanje kontinentov.

Delno lahko potrdim začetno pričakovanje, da bodo podobnosti med vizualizacijo udeležencev in dejansko vizualizacijo precejšnje. Udeleženci so največ podobnosti ugotovili v vizualizaciji Filipinov, ki so središče dogajanja, in ponazoritev, s katerih kontinentov so prihajali tviti. Najbolj očitne so večini bile podobnosti v geografskem prikazu sveta in označbi središča. Pomemben element, skupen vizualizacijam udeležencev in dejanski vizualizaciji, je bila uporaba barv za razločevanje predvsem posameznih kontinentov sveta.

Kot največje razlike med udeležencevo ter dejansko vizualizacijo so udeleženci navajali:

- dejanska vizualizacije je natančnejša;
- udeleženci vizualizirajo tudi nekatere simbole Twitterja, česar pri dejanski vizualizaciji ni;
- dejanska vizualizacija prikazuje tudi jakost poslanih tvitov;
- dejanska vizualizacija natančneje nakaže, iz katerih področij prihajajo tviti;
- udeleženec z meteorološkim znakom vizualizira tudi tajfun;
- vizualizacije udeležencev prikazujejo tudi simbole pomoči.

Ugotovitve udeležencev glede razlik med njihovo in dejansko vizualizacijo ter pomanjkljivosti dejanske vizualizacije nakazujejo, da ima dejanska vizualizacija kar nekaj pomanjkljivosti oziroma nerazumljivosti, glede na to, kaj naj bi vizualizirala. Udeleženci so

se precej osredotočali na nerazumljive elemente oziroma manjkajoče elemente, ki jih niso našli, glede na to, kar naj bi vizualizacija sporočala. Ko so povezali opis vizualizacijo in samo vizualizacijo, jih je najbolj presenetilo oziroma zmedlo, da vizualizacija z ničimer ni prikazovala, da gre za Twitterj.

Da je dejanska vizualizacija natančnejša in da so to udeleženci navajali kot eno izmed glavnih razlik med njihovo in avtorjevo vizualizacijo, ni presenečenje, saj je avtor imel na voljo precej več časa ter tudi drugačna orodja. Udeleženci so poudarjali tudi, da dejanska vizualizacija s pomočjo različne velikosti krogcev lepo prikaže tudi različno gostoto pošiljanja tvitov z različnih kontinentov.

Udeleženci so nadalje naštevali tudi glavne prednosti in izzive prve vizualizacije. Navedli so naslednje:

- **prednosti:**
  - natančnost vizualizacije,
  - ponazoritev gostote tvitov,
  - vizualizacija na nekakšen način prikaže tudi, kateri predeli sveta so razvitejši in uporabljajo Twitter;
- **izzivi:**
  - črte oziroma povezave so nerazumljive,
  - nobeden element vizualizacije ne prikaže sporočilnosti oziroma tematike,
  - legenda je precej nesmiselna in nima dodane vrednosti,
  - vizualizacija dobi smisel šele z naslovom in dodatnim opisom,
  - prevelika nakopičenost in barvitost povzroča zmedo in lahko zavaja.

Kot največje prednosti prve vizualizacije so udeleženci izpostavljali predvsem natančnost, pri čemer so imeli v mislih natančnost vizualizacije držav, celo mest, iz katerih so prihajali tviti. Prepričal jih je tudi prikaz ponazoritve gostote tvitov z večjimi ali manjšimi krogci na začetku povezav. Udeleženci so ob videni prvi vizualizaciji razmišljali tudi širše, namreč da nam vizualizacija na nek način prikaže tudi razvitost različnih delov sveta. Sklepali so, da Twitterja v revnejših predelih sveta ne poznajo in da ga ne uporabljajo. S tem moram ovreči svojo začetno postavljano domnevo, da bodo udeleženci kot prednost navajali tudi uporabo barv.

Izzivi oziroma pomanjkljivosti dejanske vizualizacije se skladajo z večino odgovorov glede razlik med njihovo in dejansko vizualizacijo. Udeleženci so imeli največ težav z nejasno uporabo elementov, kot so črte, barve in nejasnosti, ki jo te povzročajo. S tem lahko potrdim tudi začetno domnevo, da bodo udeleženci kot najmanj razumljive elemente navajali legendo in veliko število povezav.

4.1.2 Analiza rezultatov skupine s ključnimi besedami

Za razliko od vizualizacijske skupine je druga skupina najprej dobila vpogled v dejansko vizualizacijo brez dodatnega opisa in nato morala s pomočjo podajanja ključnih besed ugotoviti tematiko in naslov vizualizacije. Pred začetkom te naloge sem oblikovala kriterij ključnih besed, ki sem ga nato uporabljala za primerjavo odgovorov udeležencev in oblikovanje sklepov. Njihove odgovore oziroma ključne besede o tem, kaj predstavlja vizualizacija, sem nato umestila v določene kategorije ključnih besed. V Tabeli 6 predstavljam rezultate ujemanja ključnih besed udeležencev z vnaprej podanim kriterijem ključnih besed, prav tako dejanski čas, ki so ga udeleženci porabili za to nalogo, za katero so imeli na voljo pet minut, in naslov vizualizacije, ki so ga podali.

Tabela 6: Rezultati ujemanja udeležencevih ključnih besed z vnaprej postavljenim kriterijem pri prvi vizualizaciji

Udeleženec **Kriterij ključnih besed**		**6**	**7**	**8**	**10**	**9**
Obstoj središča	Središče	✓*	✓	✓	✓	
	Ključna točka					
	Različni deli/kontinenti sveta		✓			
Povezave	Povezanost središča z lokacijami po svetu			✓	✓	✓
	Povezave	✓			✓	✓
Kaj je središče	Filipini			✓		
	Azija	✓		✓		✓
	Država		✓			✓
	Kontinent	✓				✓
Kaj je povezava	Komunikacija z določeno vsebino po twitterju					
	Twitter					
	Pretok informacij			✓	✓	✓
	Družbeno omrežje					
Kaj je dogodek	Dogodek svetovnih razsežnosti	✓		✓	✓	
Naslov vizualizacije		Globalizacija pretoka informacij	Sesutje računalniškega omrežja	Katastrofa glede masovnih podatkov	Naravna nesreča in komunikacija po nesreči	Količina klikov
Dejanski čas (mm: ss)		04:40	04:35	4:00	05:00	03:15

Legenda: * ✓: udeležencева ključna beseda se ujema s ključno besedo, ki sem jo na začetku postavila kot kriterij.

Na osnovi rezultatov, podanih v Tabeli 6, vidim, da so štirje od petih udeležencev kot ključno besedo, kaj prikazuje vizualizacija, navajali obstoj središča celotnega dogajanja. Prav tako je večina udeležencev razbrala, da vizualizacija prikazuje precejšnje število povezav in da je središče povezano z različnimi kontinenti, državami in predeli sveta.

Udeleženci niso mogli povsem razvozlati, kaj predstavlja množica prikazanih povezav in zakaj obstoj središča. Nihče od udeležencev iz vizualizacije ni razbral, da gre za pošiljanje tvitov ali nekaj v povezavi z uporabno družbenih omrežij. So pa trije od petih udeležencev iz vizualizacije razbrali, da prikazuje nekakšen pretok informacij.

Preden so udeleženci na osnovi svojih ključnih besed morali določiti še naslov vizualizacije, sem jim podala še nekaj dejstev in podatkov v zvezi z dejansko vizualizacijo. Kljub temu da so dodatni podatki vključevali tudi to, da vizualizacija prikazuje nekaj v povezavi z družbenimi omrežji, tega ni opaziti pri dejanskih odgovorih udeležencev glede naslova vizualizacije. Res je, da so morda začeli bolj razmišljati o pretoku informacij in o komunikaciji, vendar nihče ni omenil družbenih omrežij, kot so na primer Facebook, Instagram ali druga.

Le eden od petih udeležencev je izkoristil vseh pet minut, ki jih je imel na voljo za naštevanje ključnih besed, kar kaže na to, ali da vizualizacija ni bila dovolj bogata s prikazanimi elementi, da bi pri udeležencih vzbudila dovolj domišljije, ali pa da udeleženci še niso bili dovolj navajeni branja in razmišljanja o prikazanih elementih vizualizacije.

4.1.3 Povzetek rezultatov prve vizualizacije

S primerjavo vizualizacije udeležencev in dejanske vizualizacije se ugotovi, da udeleženci zgodbo prikazujejo precej bolj preprosto, brez velikega števila črt in povezav, ki po njihovem mnenju vnašajo preveliko zmedo. Prav tako so nekateri udeleženci na svoji vizualizaciji dobro prikazali tudi simbole Twitterja, kar so pogrešali na dejanski vizualizaciji, ki naj bi prikazovala pošiljanje tvitov. Odgovori udeležencev vizualizacijske skupine nakažejo, da dejanska vizualizacija ni najbolj jasna, glede na to kakšno sporočilnost je želela prenesti. Udeleženci so navedli tudi, da samo z opazovanjem vizualizacije, brez dodatnega opisa ne bi vedeli tematike oziroma sporočilnosti, ki jo vizualizacija želi podati. Večina udeležencev vizualizacijske skupine je precej nazorno želela prikazati ključne elemente vnaprej podanega opisa dejanske vizualizacije. To sporoča, da je udeležencem zelo pomembna jasna vizualizacija, ki ne vključuje preveč elementov, hkrati pa mora prikazovati ključne elemente, in sicer glede na zgodbo, ki jo podaja.

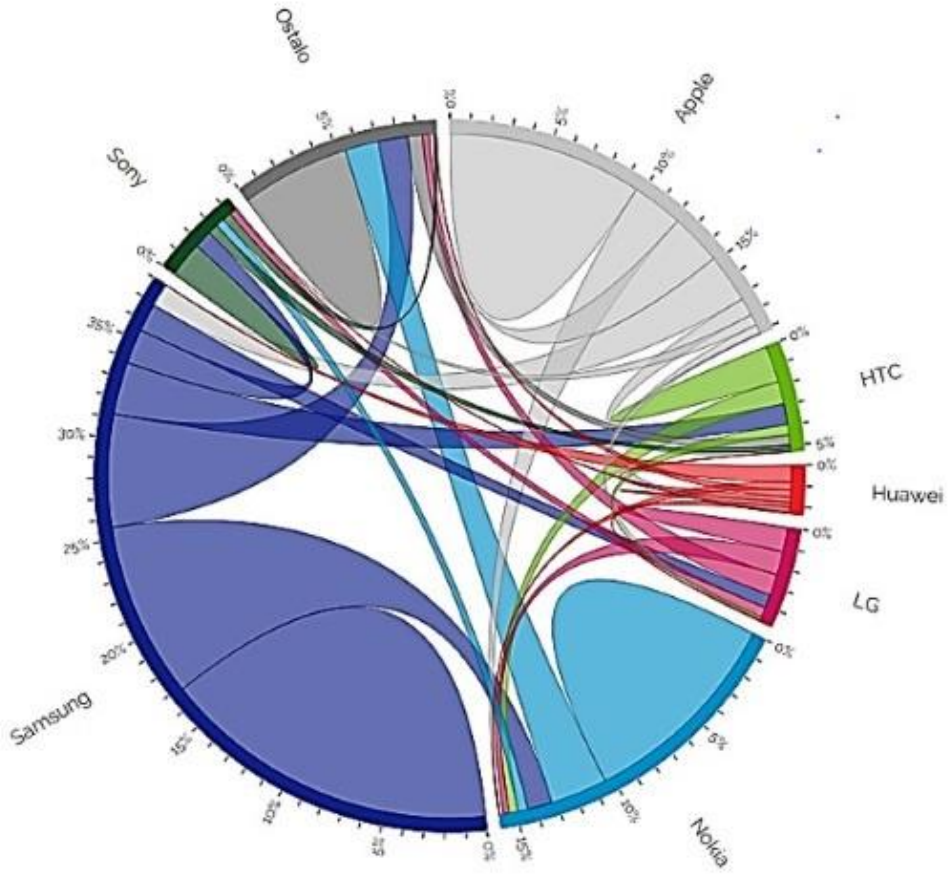
Analiza rezultatov skupine s ključnimi besedami potrjuje, da je prvi vizualizaciji manjkal kak simbol Twitterja, da bi bilo bolj razumljivo, kaj vizualizacija prikazuje in sporoča. Vendar pa je vizualizacija z vsemi vizualiziranimi črtami in povezavami, ki so se stikale v

eno točko, uspela nagovoriti udeležence, da vizualizacija prikazuje povezave oziroma povezanost s prikazano središčno točko.

4.2 Analiza rezultatov druge vizualizacije

Primer druge vizualizacije masovnih podatkov je prikazan na Sliki 5. Že pred začetkom sem pričakovala in domnevala, da bo ta udeležencem povzročala največ preglavic. Vizualizacija namreč ni tipičen prikaz grafikona ali tabele, zato lahko trdim, da je bila za večino udeležencev takšna vizualizacija nekaj novega: prikazana je v obliki kroga, znotraj katerega tetive ali loki povezujejo točke na krožnici (angl. *chord diagram*). Vizualizacija prikazuje prehajanje med blagovnimi znamkami mobilnih telefonov nizozemskih uporabnikov, temelji na anketi, izvedeni leta 2014.

Slika 4: Prehajanje med blagovnimi znamkami mobilnih telefonov med nizozemskimi uporabniki



Vir: N. Bremer, Switching behavior between phone brands of the Dutch, 2014 Edition, 2014. Ponatis z dovoljenjem.[2]

[2] Osebna komunikacija z avtorico, Nadiah Bremer po e-pošti. Dovoljenje po e-pošte pridobljeno z dnem, 22. april 2017.

4.2.1 Analiza rezultatov vizualizacijske skupine

Udeleženci so na začetku naloge dobili opis druge vizualizacije, ki se nahaja tudi spodaj. Za nalogo so imeli na voljo osem minut. Vizualizacija prehajanja med blagovnimi znamkami mobilnih telefonov med nizozemskimi uporabniki se glasi (Bremer, 2014):

> *»Skoraj vsak Nizozemec uporablja mobilni telefon in veliko uporabnikov zamenja telefon takoj, ko jim to omogoča pogodba. Ampak, ali ljudje preidejo tudi na drugo blagovno znamko in kako se ta prehod razlikuje med različnimi znamkami?*
>
> *Na to temo so izvedli anketo, potem pa so analizirali odgovore in povezave med anketirančevo trenutno in prejšnjo znamko mobilnega telefona.*
>
> *Trenutno se uporablja sedem večjih blagovnih znamk, vse ostale pa lahko združimo v eno skupino: Nokia, Huawei, HTC, Apple, LG, Samsung, Sony in ostalo. Trenutno si tržni deleži sledijo v naslednjem vrstnem redu (od največjega do najmanjšega): Samsung, Apple, Nokia, HTC, LG in Sony.*
>
> *Prikažite tudi, da je kar nekaj anketirancev prešlo iz Nokie na Samsung ter iz Samsunga na Nokio. Delež tistih, ki so preklopili iz Nokie na Samsung, je večji kot v obratni smeri.*
>
> *Nokia je izgubila veliko uporabnikov, ki so prešli na Samsung, nekaj jih je prešlo na HTC, LG in Huawei. Nekaj pa so jih tudi obdržali.*
>
> *Huawei je nova znamka, ki je vstopila na trg, nanj so preklopili največ uporabniki Samsunga, nihče od uporabnikov Appla ni naredil tega preskoka.*
>
> *HTC je pridobil nekaj Nokiinih uporabnikov in uporabnikov LG-a, nekaj jih je obdržal, še vedno pa ima majhen delež na celotnem trgu.*
>
> *Apple ni izgubil skoraj nobenega uporabnika, jih je pa kar nekaj pridobil od Samsunga, Nokie in ostalih.*
>
> *LG ima mešanico. Nekaj uporabnikov je prešlo na Samsung, Apple, HTC in Huawei, medtem ko jih je nekaj pridobil od Nokie, Sonyja in ostalih.*
>
> *Samsung je največ novih kupcev pridobil od Nokie in ostalih manjših blagovnih znamk, pridobil pa je še od vseh ostalih večjih znamk razen Appla in Huaweia.*
>
> *Sony je izgubil nekaj kupcev, ki so odšli k Samsungu, Nokii, Applu, LG-u in Huawei.«*

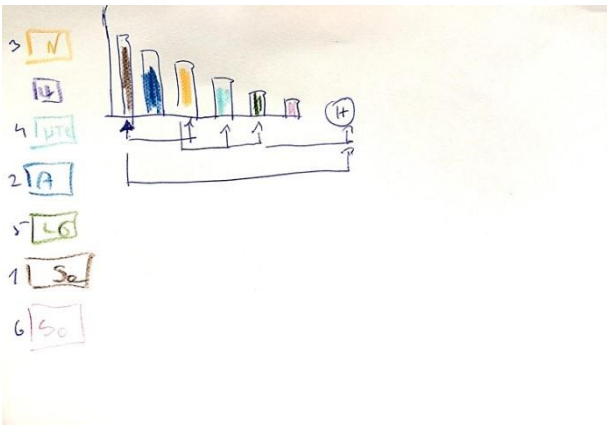
Moja pričakovanja pred začetkom nalog vizualizacijske skupine:

- na začetku bodo udeleženci zmedeni glede tega, na kakšen način deluje druga vizualizacija in kaj pomenijo različne povezave;
- odstotki, prikazani na vizualizaciji, ne bodo najbolj razumljivi;

- podobnosti med udeležencevo in dejansko vizualizacijo ne bo veliko, razen označbe velikosti tržnih deležev in vizualizacije prehodov med različnimi znamkami mobilnih telefonov;
- kot največjo razliko med svojo in dejansko vizualizacijo bodo navedli obliko vizualizacije;
- če udeleženci ne bi poznali celotne zgodbe, bi težko razložili vse elemente dejanske vizualizacije.

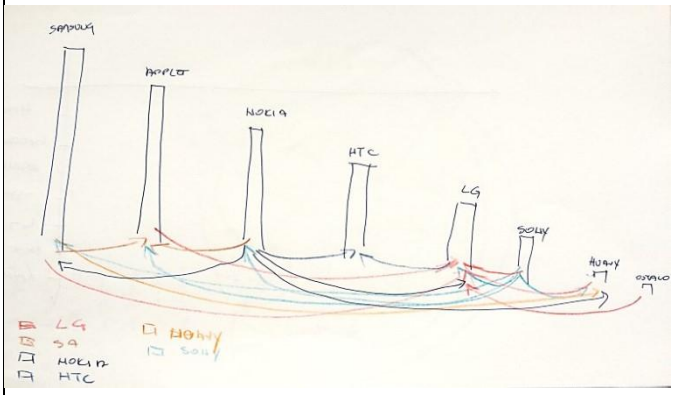
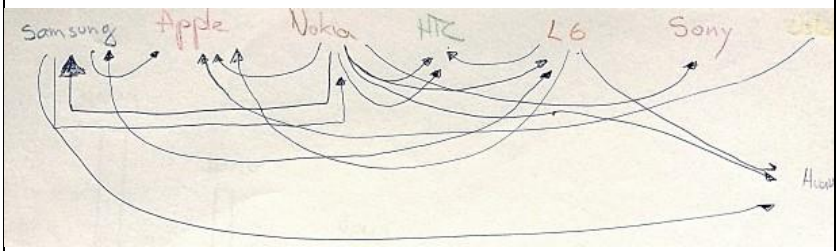
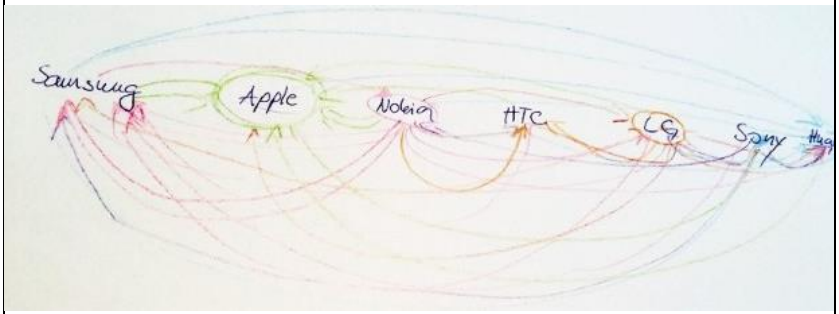
Po prebranem opisu druge vizualizacije so se udeleženci vizualizacijske skupine soočili z nalogami prvega tipa. Tabela 7 povzema rezultate, kako so se udeleženci lotili vizualiziranja, koliko časa so za to porabili (na voljo osem minut), prav tako pa tudi njihove končne vizualizacije.

Tabela 7: Rezultati poteka vizualiziranja druge vizualizacije druge skupine

Udeleženec	**Dejanski čas** (mm:ss)	**Potek vizualizacije in končna vizualizacija**
6	08:00	Udeleženec je najprej označil legendo, v kateri je vsako znamko mobilnih telefonov označil z drugo barvo. Ob legendi je oštevilčil tudi tržne deleže ob posamezni blagovni znamki. Nadaljeval je s stolpičnim grafikonom, pri čimer je največji stolpec pomenil največji tržni delež. Nato je s puščicami med stolpci ponazarjal prehode med različnimi blagovnimi znamkami, pri čemer so odebeljene pušice pomenile intenziven prehod. 

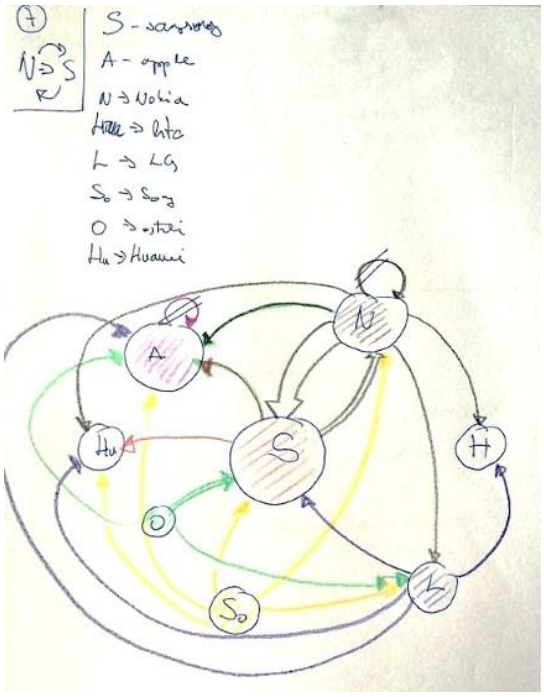
se nadaljuje

Tabela 7: Rezultati poteka vizualiziranja druge vizualizacije druge skupine (nad.)

Udeleženec	**Dejanski čas** (mm:ss)	**Potek vizualizacije in končna vizualizacija**
7	08:00	Udeleženec je vizualizacijo prikazal s stolpiči, ki so prikazovali različne znamke, pri čemer je višina stolpca pomenila velikost tržnega deleža. Barvaste puščice za različne blagovne znamke so nakazovale prehode med njimi. 
8	08:00	Udeleženec je naprej začel poimensko pisati blagovne znamke v vrstnem redu glede na tržni delež. Nato je začel risati puščice, ki so označevale prehode med blagovnimi znamkami. Debelina puščice nakazuje intenzivnost prehodov med blagovnimi znamkami. 
10	04:48	Udeleženec je najprej poimensko napisal blagovne znamke glede na velikost tržnih deležev. Nato je s puščicami vizualiziral prehode med blagovnimi znamkami. Dvojna puščica pomeni intenzivnejši prehod. 

se nadaljuje

Tabela 7: Rezultati poteka vizualiziranja druge vizualizacije druge skupine (nad.)

Udeleženec	**Dejanski čas** (mm:ss)	**Potek vizualizacije in končna vizualizacija**
9	08:00	Udeleženec je začel risati krogce v obliki kroga. Krogci, ki so vizualizirali različne blagovne znamke, so bili različnih velikosti, kar je predstavljalo različno velikost tržnih deležev. Na koncu je s puščicami različnih debelin začel nakazovati prehode med blagovnimi znamkami. 

Iz Tabele 7 razberem, da so štirje od petih udeležencev porabili vseh osem minut za vizualizacijo. Drži, da so morali prebrati tudi nekaj besedila o opisu vizualizacije, a vendar njihov čas kaže, da se niso uspeli takoj lotiti vizualiziranja oziroma si zamisliti, kako bi prikazali celotno zgodbo.

Vsi udeleženci so konkretno vizualizirali različno velikost tržnih deležev in označili različnost prehodov uporabnikov med blagovnimi znamkami mobilnih telefonov. Prav tako so vsi uporabili tudi barve za prikaz različnih povezav oziroma prehodov med blagovnimi znamkami.

Udeleženci so kot podobnosti med njihovo in dejansko vizualizacijo izpostavili:

- prikaz prehoda z ene blagovne znamke k drugi;
- izpis blagovnih znamk;
- uporaba različnih barv za različne blagovne znamke;
- prikaz tržnega deleža;
- vizualizacija s pomočjo krogcev oziroma kroga;

- na obeh vizualizacijah se prepozna, da je Samsung pridobil največ uporabnikov.

Zgoraj povzeti odgovori udeležencev pričajo o tem, da so udeleženci večinoma našli podobnosti na osnovni ravni elementov vizualizacije, kot so uporaba barv, prikaz deležev in prestopov. S tem lahko potrdim svoje pričakovanje, da bodo podobnosti v označbi velikosti tržnih deležev in prikazu prehodov med različnimi znamkami.

Kot najopaznejše razlike med njihovo in dejansko vizualizacijo so udeleženci navajali:

- dejanska vizualizacija je prikazana v obliki kroga, vizualizacija udeležencev pa v obliki grafikonov;
- dejanska vizualizacija še kvalitativno predstavi prestope;
- boljša preglednost dejanske vizualizacije;
- dejanska vizualizacija nima legende.

Pričakovala sem, da odstotki, prikazani v vizualizaciji, udeležencem ne bodo najbolj razumljivi. Do neke mere lahko to potrdim, saj so udeleženci, ko so videli dejansko vizualizacijo, takoj opazili tudi odstotke, medtem ko na svojih vizualizacijah tega niso označili.

Udeleženci so kot največjo razliko navedli obliko vizualizacije. Dva od petih udeležencev sta za vizualizacijo uporabila stolpčni diagram, dva udeleženca sta vizualizacijo naredila s pomočjo puščic in imen blagovnih znamk, eden udeleženec pa je naredil vizualizacijo v obliki kroga s pomočjo povezav in različno velikih krogcev, vsak je prikazoval drugo blagovno znamko.

Udeleženci so ob vprašanju o prednostih in izzivih druge vizualizacije navajali:

- **prednosti:**
  - prikaže veliko povezav in prestopov,
  - pregledno in nazorno prikaže tržne deleže blagovnih znamk,
  - natančnejša vizualizacija od udeležencevih;
- **izzivi:**
  - prestopi med blagovnimi znamkami so prikazani precej zapleteno,
  - ne moreš takoj jasno razbrati, kako so uporabniki prehajali med blagovnimi znamkami,
  - vizualizacija je težko razumljiva, če ne poznaš ozadja.

Večina udeležencev navaja, da bi težko ugotovili, kaj natanko predstavlja vizualizacija, če ne bi poznali celotne zgodbe, lahko pa bi sklepali, da govori o neki povezanosti med znamkami mobilnih telefonov. Kljub temu se iz odgovorov udeležencev razbere, da sama

vizualizacija prikaže vsaj en ključni element, kot je obstoj povezanosti med različnimi znamkami mobilnih telefonov. Tako lahko potrdim svoje predvidevanje, da bi brez celotne zgodbe težko ugotovili glavno temo prikaza.

4.2.2 Analiza rezultatov skupine s ključnimi besedami

V Tabeli 8 predstavljam rezultate ujemanja udeleženčevih ključnih besed z vnaprej podanim kriterijem ključnih besed, dejanski čas, ki so ga udeleženci porabili za to nalogo (na voljo so imeli pet minut), ter naslov vizualizacije, ki so ga podali.

Tabela 8: Rezultati ujemanja udeleženčevih ključnih besed z vnaprej postavljenim kriterijem pri drugi vizualizaciji

Udeleženec **Kriterij ključnih besed**		**1**	**2**	**3**	**4**	**5**
Ključna tema	Prehodi med različnimi znamkami mobilnih telefonov					
	Večje znamke mobilnih telefonov	✓*		✓	✓	✓
	Mobilni telefoni	✓	✓	✓	✓	✓
Povezave	Povezujejo različne znamke mobilnih telefonov	✓	✓		✓	✓
	Povezava		✓		✓	✓
	Razmerje			✓		
Pomen barve povezav	Specifična blagovna znamka	✓			✓	
	Pomen velikosti povezav	✓	✓	✓	✓	
	Število uporabnikov, ki so preklopili med blagovno znamko					
Kaj je povezava	Preklop med blagovnimi znamkami					
Kaj je dogodek	Spreminjanje uporabe blagovnih znamk mobilnih telefonov					
Naslov vizualizacije		Povezave med proizvajalci mobilnih telefonov na globalni ravni	Povezave med mobilnimi podjetji	Razmerje deležev posameznih proizvajalcev	Odstotek uporabnikov, ki so zvesti določeni znamki	Virtualni svet
Dejanski čas (mm:ss)		05:00	02:40	02:00	04:38	04:56

Legenda: * ✓: udeleženčeva ključna beseda se ujema s ključno besedo, ki sem jo na začetku postavila kot kriterij.

Iz rezultatov v Tabeli 8 razberem, da je večina udeležencev pri tej nalogi glede na zgornjo razporeditev kriterija ključnih besed podala enak ali zelo podoben odgovor vsaj glede

ključne teme, povezav in pomena barve povezav. Pričakovano so vsi udeleženci iz vizualizacije razbrali, da ta prikazuje nekaj v povezavi z mobilnimi telefoni, večina je opazila tudi, da gre pri vizualizaciji za večje blagovne znamke mobilnih telefonov. Do teh sklepov so prišli kaj kmalu ob začetku te naloge. Navajanje ključnih besed, kaj prikazujejo vse povezave ter prepletanja barv in črt na vizualizaciji, pa je bilo za udeležence težje. Pri tem so udeleženci navajali, da povezave in prepletanja lahko pomenijo, da uporabniki uporabljajo dve znamki mobilnih telefonov, nakupne navade potrošnikov, produktni portfelj naštetih blagovnih znamk, tekmovanje, soodvisnost, prodajo telefonov do pripadnosti blagovnim znamkam oziroma zadovoljstvo kupcev.

Odgovori udeležencev, kakšen naslov bi dali pokazani vizualizaciji na osnovi svojih ključnih besed, so si precej podobni. Večinoma v naslovu namreč omenjajo povezave med znamkami mobilnih telefonov, eden udeleženec pa se je zelo približal dejanski sporočilnosti vizualizacije, ko je kot naslov omenil pripadnost blagovni znamki mobilnih telefonov.

4.2.3 Povzetek rezultatov druge vizualizacije

Pri drugi vizualizaciji sta tako vizualizacijska skupina kot skupina s ključnimi besedami imeli največ težav pri prepoznavanju točne teme prikaza. Analiza rezultatov razkrije, da udeleženci niso bili vajeni na prikaze prehodov oziroma premikov v obliki diagrama, ki ga prikazuje druga vizualizacija.

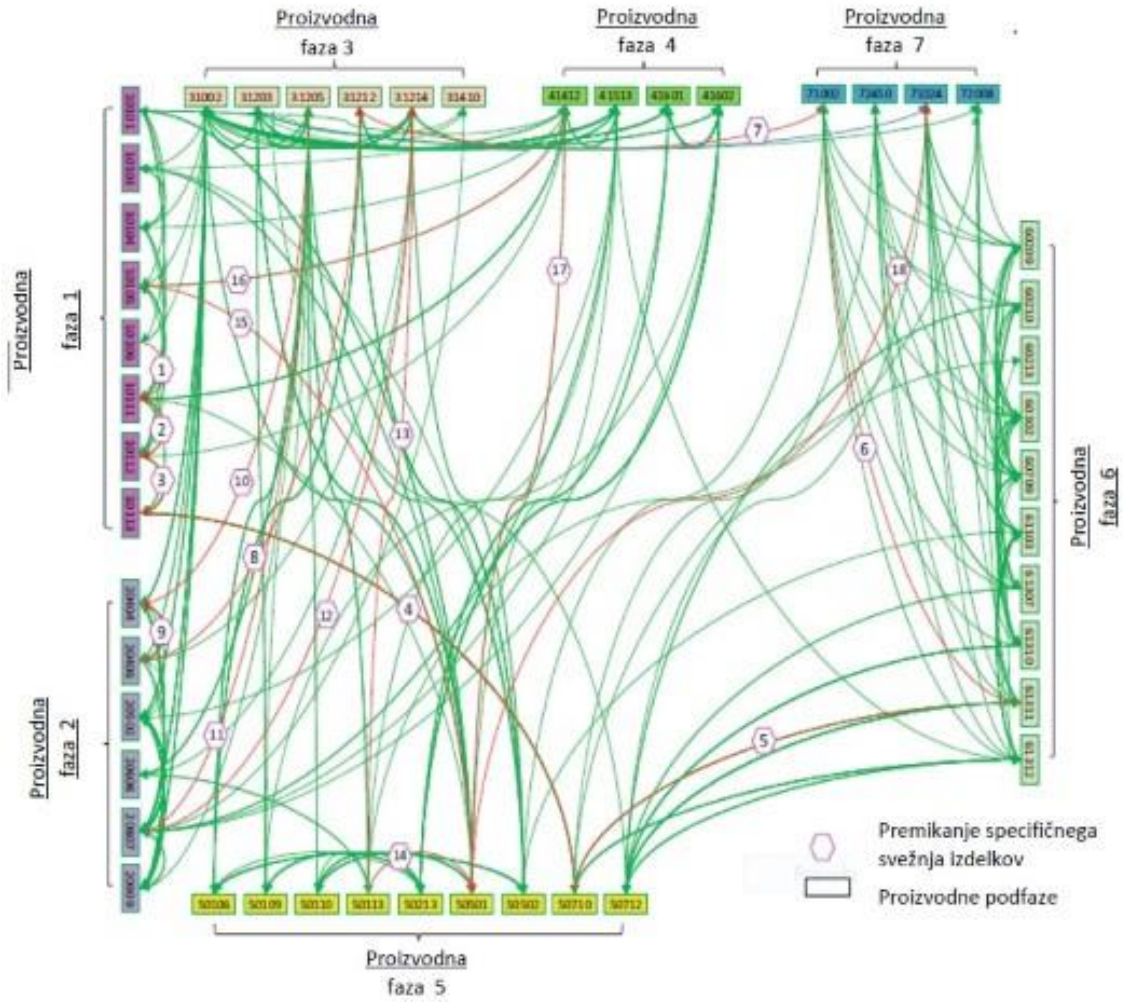
Vizualizacijska skupina, ki je pred sabo imela opis, kaj vizualizacija prikazuje, je večinoma uporabila tradicionalne načine prikaza, torej preproste puščice in črte, ki so prehajale od ene do druge blagovne znamke. Njihova uporaba barv pri vizualiziranju potrjuje, da uporabniki vizualizacij za razumevanje osnovne teme prikaza potrebujejo tudi ta element, za lažje razumevanje sporočilnosti.

Analiza rezultatov udeležencev skupine s ključnimi besedami nadalje potrjuje, da je udeležencem način prikaza druge vizualizacije bil novost. Sicer pa so zaradi uporabljenih elementov, kot so imena blagovnih znamk, uspeli razbrati, da se nahajamo v področju mobilne telefonije in da so blagovne znamke povezane med sabo. Udeležencem pa je bil največji izziv določiti, kaj predstavljajo te povezave, kar razkrivajo tudi njihovi odgovori, kakšen naslov bi dali prikazani vizualizaciji. Ob tem so udeleženci navajali sicer širok spekter odgovorov in s tem pokazali, kako različno lahko interpretiramo vizualizacijo, če ne poznamo vseh podatkov.

4.3 Analiza rezultatov tretje vizualizacije

Tretja vizualizacija, na Sliki 6, prikazuje logistične poti s pomočjo brezžičnih radijskih valov. Prav tako kot ostali dve vizualizaciji na nek način prikazuje pot oziroma nekakšno povezanost.

Slika 5: Prikaz logističnih poti s pomočjo brezžičnih radijskih valov



Vir: R.Y. Zhong, S. Lan, C. Xu, Q. Dai in G. Q. Huang, Visualization of RFID-enabled shopfloor logistics Big Data in Cloud Manufacturing, 2016, str. 12. Ponatis z dovoljenjem.[3]

4.3.1 Analiza rezultatov vizualizacijske skupine

Ta skupina udeležencev je ponovno dobila opis dejanske vizualizacije, ki ga navajam v nadaljevanju. Dejanske vizualizacije pa udeleženci niso videli do trenutka, ko so vizualizirali svojo, temelječo na opisani zgodbi (Zhong et al., 2016):

> *»Podjetje Huaiji s Kitajske je specializirano v proizvodnji ventilov motorja, pri čemer uporablja najnovejše tehnologije. Od leta 2003 naprej so svoje napore usmerili tudi v izboljšanje notranjih procesov. Da bi pridobili podatke o trenutnem stanju in jih uporabili tudi za izboljšanje, so zgradili tehnologije brezžičnih radijskih valov. V primeru logističnih operacij v enem proizvodnem oddelku to pomeni, da so na različne korake proizvodnje in premikanja produkta iz ene v drugo fazo proizvodnje namestili brezžične vmesnike. Ti zaznane informacije pošiljajo direktno v bazo, kjer se ti zbirajo.«*

[3] Komunikacija z založnikom Springer po e-pošti. Dovoljenje pridobljeno z dnem 8. oktober 2016.

»V proizvodnem obratu je sedem različnih faz, prek katerih lahko gredo izdelki. Vsaka izmed sedmih proizvodnih faz ima še podfaze, v katerih potekajo različna dela proizvodnje izdelka. Število podfaz v različnih fazah je različno:

- *faza 1: 8 podfaz*
- *faza 2: 6 podfaz*
- *faza 3: 6 podfaz*
- *faza 4: 4 podfaze*
- *faza 5: 9 podfaz*
- *faza 6: 10 podfaz*
- *faza 7: 4 podfaze*

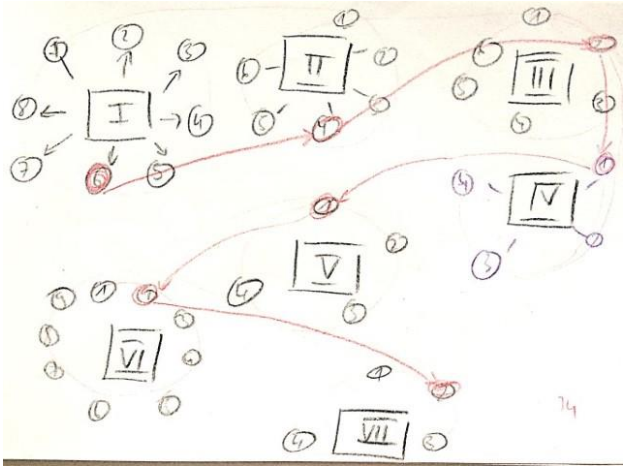
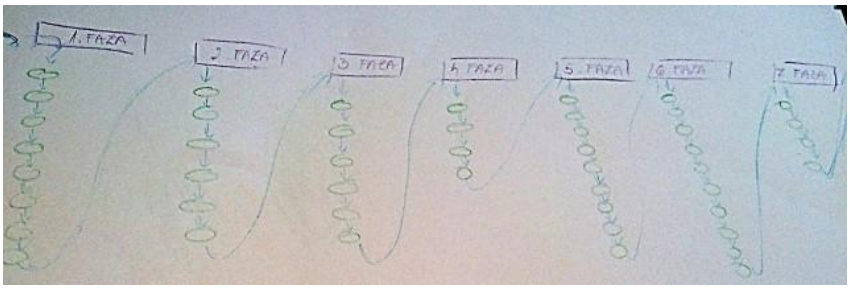
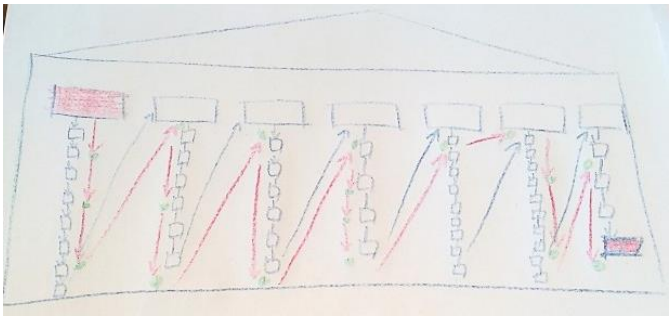
Zamislite si proizvodnjo enega paketa izdelkov. Želimo prikazati, kako se ta paket izdelkov tekom proizvodnje premika med različni fazami in podfazami. Predstavljajmo si, da gre ta specifični paket izdelkov tekom proizvodnje prek 18 podfaz, zato se mora 18-krat premakniti, da na koncu dobimo končni paket izdelkov.«

Moja pričakovanja pred začetkom naloge vizualizacijske skupine:

- podobnosti med udeležencevo in dejansko vizualizacijo bodo precejšnje, saj bodo udeleženci prikazali veliko povezav med proizvodnimi in vmesnimi postajami ter tudi osemnajst premikov, kot omenjeno v opisu dejanske vizualizacije;
- kot izzive in težje razumljivo bodo navedli veliko število povezav, ki jih prikazujeta vizualizacija in njihovo prepletanje;
- zaradi velikega prepletanja povezav na vizualizaciji in relativno skope legende bi udeleženci težko ugotovili, kaj prikazuje vizualizacija, če ne bi poznali dodatne zgodbe.

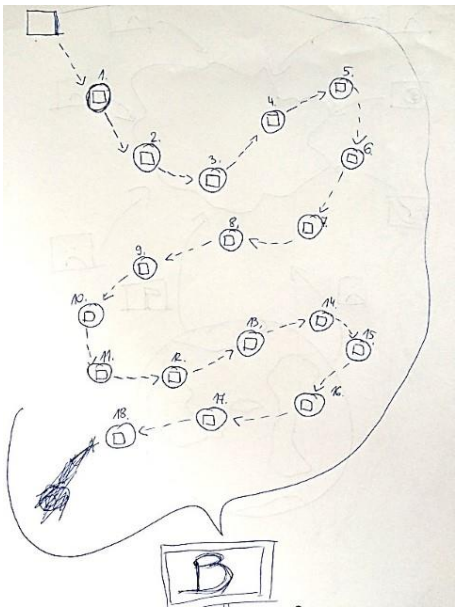
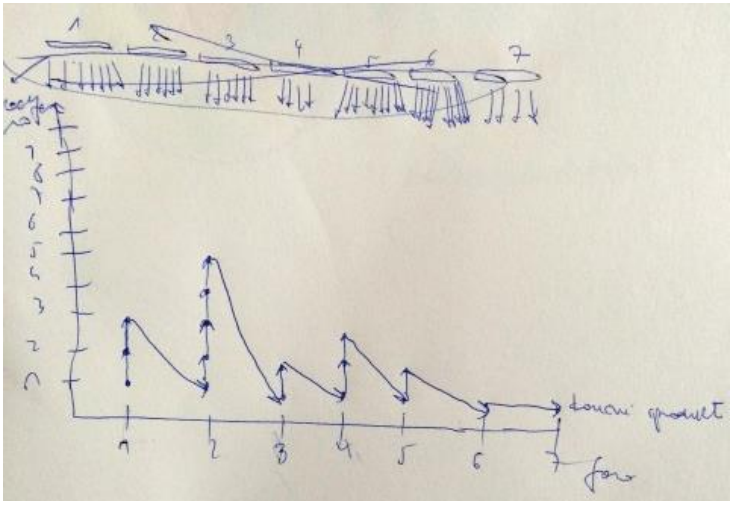
V Tabeli 9 je povzetek rezultatov, kako so se udeleženci lotili svoje vizualizacije, saj to poda pogled v njihovo razmišljanje v času ustvarjanja. Prikazujem tudi njihove končne vizualizacije in podam dejanski čas, ki so ga porabili, na voljo so imeli pet minut.

Tabela 9: Rezultati poteka vizualiziranja tretje vizualizacije vizualizacijske skupine

Udeleženec	**Dejanski čas** (mm:ss)	**Potek vizualizacije in končna vizualizacija**
1	05:00	Udeleženec je najprej razmišljal, kako vizualizirati faze in podfaze. Nato je faze prikazal s kvadratki, podfaze pa s krogci. Na koncu si je še zamislil pot specifičnega svežnja izdelkov in jo vizualiziral s posebno barvo in črto skozi krogce. 
4	04:40	Udeleženec je najprej prikazal faze v obliki kvadratkov, pod vsakim izmed njih pa še krogce, ki so simbolizirali podfaze. S puščico na začetku je udeleženec nakazal vhod svežnja izdelkov v prvo proizvodno fazo. S puščicami je nato vizualiziral pot svežnja izdelkov skozi vse vizualizirane faze in podfaze. 
5	05:00	Udeleženec je najprej z ikonami vizualiziral proizvodni obrat. Znotraj proizvodnega obrata je nato vizualiziral kvadrate, ki so predstavljali faze, pod njimi pa kvadratke, ki so vizualizirali podfaze. S puščicami je nato povezal vse faze in podfaze. Nadaljeval je z drugačno barvo in z njo označil še pot specifičnega svežnja izdelkov. Na koncu je s krogci vizualiziral še brezžične vmesnike na vsak premik svežnja izdelkov. 

se nadaljuje

Tabela 9: Rezultati poteka vizualiziranja tretje vizualizacije vizualizacijske skupine (nad.)

Udeleženec	**Dejanski čas** (mm:ss)	**Potek vizualizacije in končna vizualizacija**
2	04:31	Na začetku je udeleženec imel nekaj vprašanj predvsem glede poti specifičnega svežnja izdelkov. Udeleženec je najprej označil kvadrat, ki je predstavljal paket izdelkov, ki je vstopal v proizvodni proces. Označil je podfaze in jih s črtkanimi puščicami, ki so predstavljale brezžične vmesnike, povezal med sabo. Na koncu je vizualiziral tudi ikono računalnika, kamor se na koncu zbirajo informacije iz proizvodnega procesa. 
3	04:39	Udeleženec je najprej začel z vizualiziranjem faz v obliki kvadratkov. Iz vsakega izmed njih je izpeljal puščice, vsaka predstavlja eno podfazo. Nato se je lotil drugačne strategije. V grafikon je na abscisni osi označeval faze, na ordinatni osi pa podfaze. Sledeč opisu dejanske vizualizacije je s pikami označil podfaze in jih nato primerno na koncu povezal in na tak način vizualiziral gibanje specifičnega paketa izdelkov. 

Iz rezultatov v Tabeli 9 je vidno, da je večina udeležencev zelo sledila predhodnemu opisu zgodbe glede dejanske vizualizacije, ki sem jim ga podala. Glede na to, da so imeli na voljo za svojo vizualizacijo pet minut, so čas skoraj vsi udeleženci tudi do konca izkoristili. Vsi udeleženci so namreč porabili najmanj štiri minute in trideset sekund ali več.

Trije od petih udeležencev so se vizualiziranja lotili s prikazom faz in podfaz v obliki kvadratov ali krogcev, ki so jih nato s puščicami povezovali med sabo, s čimer so prikazovali potovanje svežnja izdelkov. Eden izmed teh treh udeležencev je razmišljal še širše in dejansko prikazal oziroma vsaj poizkušal nakazati, da gre za zbiranje podatkov med proizvodnim procesom s pomočjo brezžičnih vmesnikov v neko skupno računalniško bazo.

Eden udeleženec je prikazal le podfaze po vrstnem redu in jih med sabo sistematično povezal. Tudi ta udeleženec je razmišljal širše in poizkušal nakazati stekanje podatkov v računalniško bazo. Eden izmed petih udeležencev pa je prikaz naredil v obliki linijskega grafikona, pri čemer je na različnih oseh označil faze in podfaze, ki jih je nato po vrstnem redu povezal skupaj.

Glede na zgoraj podano analizo rezultatov izpostavljam udeleženca z naravoslovno izobrazbo, ki se je lotil vizualizacije bolj s tehničnega vidika, saj je njegova vizualizacija v obliki linijskega grafikona.

Udeleženci so kot podobnosti med njihovo in dejansko vizualizacijo izpostavili:

- podoben shematski prikaz;
- s številkami označene podfaze;
- razdeljenost in označba faz ter podfaz;
- povezave;
- prikaz poteka poti določenega svežnja izdelkov.

Vizualizacije udeležencev in dejanska vizualizacija so si bile podobne, če upoštevam, da so vse vključevale vizualizacijo proizvodnih faz in podfaz ter premik specifičnega svežnja izdelkov, s čimer lahko potrdim svojo prvo postavljeno hipotezo.

Udeleženci so med njihovo in dejansko vizualizacijo navajali naslednje najopaznejše razlike:

- dejanska vizualizacija natančneje prikazuje prehode;
- dejanska vizualizacija je težje razumljiva zaradi veliko vizualiziranih prehodov;
- različna vizualizacija poteka valov;
- dejanska vizualizacija ne prikazuje, da se informacije stekajo v bazo podatkov;
- dejanska vizualizacija podfaze označuje s kodami;
- na dejanski vizualizaciji ni označen vhod izdelka.

Udeleženci so imeli pri dejanski vizualizaciji največ preglavic z njeno kompleksnostjo, saj zaradi vseh prikazanih povezav in njihovega prepletanja na prvi pogled ni mohoče razbrati bistva. Tudi ta povzetek se sklada z začetnimi predvidevanji o tem, kaj bo udeležencem povzročalo največ težav.

Udeleženci so na vprašanje o prednostih in izzivih druge vizualizacije navajali:

- **prednosti:**
  - dober prikaz kompleksnosti opisanega,
  - natančnejša vizualizacija, saj prikaže celotno kompleksnost proizvodnega obrata, v katerem so nameščeni brezžični vmesniki,
  - jasna slika premika premikov specifičnega svežnja izdelkov;
- **izzivi:**
  - prevelika zapletenost oziroma kompleksnost vseh črt in povezav,
  - preveč črt in povezav,
  - manjša preglednost, kar onemogoča večje slike celotne vizualizacije.

Kot največji izziv tretje vizualizacije so udeleženci večinoma navajali preveliko zapletenost oziroma kompleksnost vseh črt in povezav, ki jih je nekoliko preveč. Navedli so tudi manjšo preglednost, ki otežuje prikaz jasnejše slike.

Odgovori udeležencev so se precej razlikovali pri vprašanju, ali bi samo iz vizualizacije, brez dodatnega opisa, lahko ugotovili vsebino zgodbe. Nekateri so navedli, da bi s pomočjo legende to lahko ugotovili, drugi so poudarili, da je vizualizacija prezapletena, da bi lahko brez dodatnega opisa izvedeli kaj več. Večina udeležencev je dodatno navedla, da bi iz vizualizacije lahko vseeno predvidevali, da gre za okolje, v katerem se dogaja proizvodni proces, te vlogo imajo enote in podenote. Glede na to lahko le delno potrdim začetno domnevo, da bodo udeleženci težko ugotovili, kaj prikazuje vizualizacija brez dodatne zgodbe.

4.3.2 Analiza rezultatov skupine s ključnimi besedami

V Tabeli 10 predstavljam, kako so se odgovori udeležencev skladali z mojim vnaprej predpostavljenim kriterijem ključnih besed. Prikazujem tudi dejanski čas, ki so ga udeleženci porabili, na voljo so imeli pet minut. Poleg tega sem beležila, kakšen naslov vizualizacije bi dali udeleženci glede na videno, in njihove ključne besede.

Tabela 10: Rezultati ujemanja udeležencevih ključnih besed z vnaprej postavljenim kriterijem pri drugi vizualizaciji

Udeleženec **Kriterij ključnih besed**	**7**	**8**	**9**	**10**	**6**
Proizvodnja	✓*	✓	✓	✓	
Povezave	✓	✓	✓	✓	✓
Faze				✓	✓
Podfaze					✓
Proizvodne postaje			✓		✓
Proizvodne faze in podfaze				✓	✓
Poveza med različnimi postajami	✓		✓	✓	✓
Logistične povezave			✓	✓	
Logistična pot			✓	✓	
18 povezav		✓			
7 faz					
Naslov vizualizacije	Elektrotehnika v kmetijstvu	Tehnološki proces	Vizualizacija povezav med postajami elektro procesa	Vizualizacija logističnih poti	Značilnosti premikanja materialov
Dejanski čas	02:30	03:43	03:14	03:00	04:10

Legenda: * ✓: udeležencева ključna beseda se ujema s ključno besedo, ki sem jo na začetku postavila kot kriterij

Iz rezultatov v Tabeli 10 razberem, da so štirje od petih udeležencev iz dejanske vizualizacije sklepali, da se nahajamo v proizvodnji oziroma v proizvodnem obratu. Vsi udeleženci so kot eno izmed prvih ključnih besed navajali povezave. Samo dva od petih udeležencev sta kot ključno besedo navedla faze, samo eden pa podfaze, čeprav je to bilo podano tudi v legendi ob vizualizaciji. Dva od petih udeležencev sta kot ključni besedi navedla logistične povezave in logistično pot.

Ostale navedene ključne besede udeležencev, ki jih nisem mogla uvrstiti v svojo kategorizacijo, dajejo vpogled v široko razmišljanje udeležencev ob tej vizualizaciji. Nekatere izmed njim so bile na primer medsebojna pomoč oziroma sodelovanje v proizvodnji, količina narejenih izdelkov na osebo, učinkovitost proizvodnje, prikaz načrtovanega in dejanskega poteka proizvodnje, problem, različni procesi pri izdelavi izdelka. To potrjuje, da so udeleženci razmišljali širše, a vendar še vedno ostajali blizu področja, ki ga dejanska vizualizacija prikazuje.

Odgovori udeležencev glede naslova vizualizacije so si precej različni. Pa vendar so vsi v okviru področja tehničnih ved, ali govorijo o materialih, elektrotehniki ali pa o logističnih poteh. Eden izmed udeležencev se je s svojim naslovom zelo približal dejanskemu naslovu vizualizacije, saj je za naslov navedel vizualizacijo logističnih poti.

Glede na to, da so imeli udeleženci za naštevanje ključnih besed na voljo pet minut, je iz tabele zgoraj vidno, da nihče od udeležencev ni docela izkoristil danega časa. V povprečju so udeleženci za to nalogo porabili tri minute.

4.3.3 Povzetek rezultatov tretje vizualizacije

Tretja vizualizacija je bila zadnja izmed dejanskih vizualizacij masovnih podatkov, ki prikazujejo poti, ki sem jo uporabila v empiričnemu delu magistrskega dela. Udeleženci so torej že imeli nekaj izkušenj s tem, kako brati vizualizacije in kakšne naloge jih čakajo.

Pri tretji vizualizaciji, ki je prikazovala logistične poti s pomočjo brezžičnih radijskih valov, analiza odgovorov udeležencev obeh skupin priča o tem, da je ta vizualizacija bila udeležencem z vidika glavne teme prikaza sicer razumljiva. Pa vendar je obema skupinama udeležencev težave povzročala kompleksnost vseh prikazanih povezav; ta je morda celo zavajala in oteževala preprost prikaz želene teme. Dejanska vizualizacija vključuje preveč elementov in je preveč prepletena, da bi nekdo lahko hitro razbral glavno temo prikaza.

4.4 Primerjava prikazov in končne ugotovitve

Zadnje vprašanje, ki sem ga zastavila vsem udeležencem, se je nanašalo na to, katera izmed vizualizacij se jim je zdela najbolj razumljiva glede na jasnost prikazane teme. Rezultate njihovih odgovorov povzemam v Tabeli 11.

Tabela 11: Razvrstitev vizualizacij glede na njihovo razumljivost

Udeleženec	Starostni razred	Stopnja in smer izobrazbe	Najmanj razumljiva vizualizacija	Srednje razumljiva vizualizacija	Najbolj razumljiva Vizualizacija
8	Od 20 do 31	Visokošolska univerzitetna izobrazba (naravoslovna smer)	Vizualizacija 1	Vizualizacija 2	Vizualizacija 3
4	Od 20 do 31	Visokošolska univerzitetna izobrazba (družboslovna smer)	Vizualizacija 1	Vizualizacija 2	Vizualizacija 3
5	Od 20 do 31	Magisterij stroke (družboslovna smer)	Vizualizacija 1	Vizualizacija 3	Vizualizacija 2
7	Od 32 do 51	Srednja poklicna šola (naravoslovna smer)	Vizualizacija 1	Vizualizacija 3	Vizualizacija 2
2	Od 20 do 31	Visokošolska univerzitetna izobrazba (družboslovna smer)	Vizualizacija 2	Vizualizacija 3	Vizualizacija 1
10	Od 20 do 31	Magisterij stroke (družboslovna smer)	Vizualizacija 2	Vizualizacija 1	Vizualizacija 3

se nadaljuje

Tabela 11: Razvrstitev vizualizacij glede na njihovo razumljivost (nad.)

Udeleženec	Starostni razred	Stopnja in smer izobrazbe	Najmanj razumljiva vizualizacija	Srednje razumljiva vizualizacija	Najbolj razumljiva Vizualizacija
3	Od 20 do 31	Doktor znanosti (naravoslovna smer)	Vizualizacija 2	Vizualizacija 3	Vizualizacija 1
1	Od 32 do 51	Magisterij stroke (družboslovna smer)	Vizualizacija 2	Vizualizacija 3	Vizualizacija 1
9	Od 32 do 51	Doktor znanosti (naravoslovna smer)	Vizualizacija 2	Vizualizacija 1	Vizualizacija 3
6	Od 32 do 51	Doktor znanosti (družboslovna smer)	Vizualizacija 3	Vizualizacija 2	Vizualizacija 1
Skupaj vizualizacija 1			**4**	**2**	**4**
Skupaj vizualizacija 2			**5**	**3**	**2**
Skupaj vizualizacija 3			**1**	**5**	**4**

Legenda: Vizualizacija 1: prikaz aktivnosti na Twitterju po katastrofalnem tajfunu na Filipinih novembra 2013; vizualizacija 2: prehajanje med blagovnimi znamkami mobilnih telefonov med nizozemskimi uporabniki; vizualizacija 3: prikaz logističnih poti s pomočjo brezžičnih radijskih valov

Iz Tabele 11 razberem, da se odgovori udeležencev precej razlikujejo, a kljub temu iz povzetka lahko vidim, da je petim udeležencem najmanj razumljiva druga vizualizacija, sledi ji prva vizualizacija. Ob tem so udeleženci kot razlog za nerazumljivost največkrat navedli, da iz vizualizacije sporočilo ni bilo jasno razumljivo zaradi nerazumljivih povezav in razmerij, prevelike nakopičenosti elementov ter manjkajočih elementov o prikazani temi vizualizacije.

Štirje udeleženci so kot najrazumljivejšo vizualizacijo navedli prvo, štirje udeleženci pa tretjo vizualizacijo. Utemeljitev, da so kot najrazumljivejšo navedli prvo vizualizacijo, so največkrat navajali, da je najbolj logično prikazala dogodek in da je vizualizacija natančna ter najpreglednejša. Za tretjo vizualizacijo so udeleženci izbiro utemeljili z razlago, da se lahko iz vizualizacije pregledno vidi, da ta prikazuje sklope in podenote, ki so med sabo prepletene in se premikajo. V veliko pomoč navajajo tudi legendo, eden udeleženec pa vsakodnevno dela v logističnem oddelku proizvodnega obrata, zato mu je bila tematika bližja in tudi razumljivejša.

Rezultati v Tabeli 11 ne kažejo na podobnosti odgovorov udeležencev istega starostnega razreda ali stopnje in smeri izobrazbe. Ne morem sklepati, da starost, smer in stopnja izobrazbe vplivajo na udeležencevo razvrstitev glede na razumevanje vizualizacij.

Na osnovi rezultatov empiričnega dela magistrskega dela podajam ugotovitve v podporo na začetku postavljenih hipotez empiričnega dela in tiste, ki so v prid zavrnitvi hipotez. Prva hipoteza je bila, da ljudje vizualizacijo masovnih podatkov gledamo skozi tradicionalne

metode grafičnih prezentacij. K tem metodam spadajo predvsem stolpični, razsevni in linijski grafikoni ter strukturni krog. Za to hipotezo je v rezultatih veliko podpore, saj so se udeleženci vizualizacijskih skupin pri nalogah, pri katerih so morali vizualizirati opisano zgodbo, precej posluževali teh tradicionalnih metod. Tudi skupina s ključnimi besedami je imela težave z netradicionalnimi prikazi.

Druga hipoteza se je nanašala na starost. Domnevala sem namreč, da so vizualizacije hitreje razumljive mlajšim udeležencem, saj so ti bolj informirani o najnovejših trendih na področju kopice podatkov in njihove vizualizacije. Glede na rezultate empiričnega dela magistrskega dela ne morem potrditi te hipoteze. Rezultati namreč ne kažejo na to, da bi mlajši udeleženci ali bistveno hitreje ali bolje razumeli predstavljene vizualizacije.

Naslednja hipoteza je bila, da predznanje z določenega področja izboljša razumevanje vizualizacij masovnih podatkov s tega področja. Za potrditev te hipoteze v empiričnem delu nisem našla zelo prepričljive podlage. Največ mi pomagajo pri potrjevanju ali ovržbi te hipoteze rezultati skupine s ključnimi besedami, kjer najdem nekaj podpore za to hipotezo. Udeleženci skupine s ključnimi besedami, ki so imeli predznanje o tematiki, ki jo je prikazovala vizualizacija, so se bolj približali dejanskemu naslovu vizualizacije, prav tako so so njihove besede večkrat ujemale s postavljenim kriterijem ključnih besed. Pa vendar tega ne morem pripisovati izkjučno temu, da jim je pri tem pomagalo predznanje določenega področja.

Četrta hipoteza je bila, da pomanjkljiva vizualizacija masovnih podatkov lahko poda napačno sporočilnost. Tej hipotezi lahko pritrdim predvsem v primeru prve vizualizacije, vizualizacije aktivnosti na Twitterju po katastrofalnem tajfunu na Filipinih novembra leta 2013. Pri tej vizualizaciji sta namreč obe skupini udeležencev pogrešali ključni element, vizualizacijo simbola Twitterja.

Pri peti hipotezi sem domnevala, da obstaja razlika v razumevanju in hitrosti dojemanja vizualizacije, če nekdo naprej vidi vizualizacijo masovnih podatkov in poizkuša ugotoviti bistvo prikaza, ali pa na drugi strani nekdo najprej sliši zgodbo vizualizacije in šele nato dejansko vidi vizualizacijo. To lahko potrdim, saj je v empiričnem delu vizualizacijska skupina najprej videla opis vizualizacije in nato vizualizacijo. Na ta način je vizualizacijska skupina že prej lahko dobila bistvo teme prikaza. Drži, da so se udeleženci še vedno soočali s težavami dejanske vizualizacije glede določenih elementov, a vendar so vedeli, kaj naj bi vizualizacija prikazovala. Na drugi strani je skupina s ključnimi besedami težje ugotovila, kaj je glavna tema prikaza, saj udeleženci najprej niso poznali dodatnega opisa zgodbe. Iz tega lahko sklepam, da je z vidika razumljivosti glavne teme prikaza pomembno ubesediti bistvo zgodbe (npr. skozi naslov ali dodatne opombe).

SKLEP

Masovni podatki in njihova vizualizacija so trend, ki prihaja v vsakodnevna življenja vseh nas. Nista samo zasebni in javni sektor tista, ki lahko imata koristi od tega, ampak vsak izmed nas tudi pri vsakodnevnih opravilih. Prav zaradi tega sem se lotila pisanja magistrskega dela na temo analize vizualizacij masovnih podatkov, ki prikazujejo poti. Vizualiziranje poti oziroma premikov se mi še vedno zdi ena pogosteje uporabljenih tem vizualizacij, prav tako se je zagotovo vsakdo že srečal z vizualizacijo tega motiva.

Namen vizualizacije masovnih podatkov je predstaviti masovne podatke na razumljivejši in preprostejši način. S pomočjo vizualizacije moramo dobiti kolikor se le da najboljši vpogled v bistvo glavne teme prikaza, ne da bi ob tem morali poznati celotno ozadje in vse podrobnosti.

V teoretičnem delu magistrskega dela sem opredelila masovne podatke in vizualizacijo masovnih podatkov, poseben poudarek sem namenila vodilom in priporočilom za dobro vizualizacijo, kot jih najdemo v strokovni literaturi. Nato sem v empiričnem delu magistrskega dela preverjala, koliko so ta vodila upoštevana pri treh dejanskih primerih vizualizacij masovnih podatkov. S pomočjo teh treh primerov sem na majhnem raziskovalnem vzorcu proučevala, kako dobro te vizualizacije prikazujejo glavno temo prikaza. Glavni namen magistrskega dela je torej bil raziskati, koliko so vizualizacije masovnih podatkov res razumljive in ali opravičujejo svoj namen.

Glede na prebrano strokovno literaturo in rezultate, pridobljene v empiričnem delu magistrskega dela, lahko podam naslednje sklepe o na začetku dela postavljenih temeljnih hipotezah, pri čemer pa je treba upoštevati, da so pridobljeni na majhnem raziskovalnem vzorcu. Rezultati nakazujejo, da vizualizacije masovni podatkov še vedno gledamo skozi tradicionalne metode grafičnih prezentacij. To namreč sklepam na osnovi rezultatov tega, kako so udeleženci pri vizualizacijskih nalogah opisano zgodbo še vedno prikazovali večinoma z linijskimi ali stolpičnimi grafikoni ter s preprostimi ikonami, ki so jih potem povezovali med sabo glede na podani opis.

Rezultati tudi kažejo, da pomanjkljiva vizualizacija masovnih podatkov lahko da napačno sporočilnost. Temu pritrjujejo rezultati udeležencev pri vseh treh dejanskih vizualizacijah. Pomanjkljiva oziroma netočna vizualizacija lahko celo zmede uporabnika končne vizualizacije in ga odvrne od bistva prikazane teme. Pri dejanskih vizualizacijah, ki jim je manjkal kak ključni element glede na želeno glavno temo prikaza, so bili udeleženci zmedeni in niso ugotovili glavne sporočilnosti oziroma so se od nje oddaljili.

V empirični raziskavi pa nisem našla veliko podpore za potrditev hipoteze o vplivu starosti in predznanja o prikazani tematiki na razumevanje glavne sporočilnosti prikazanega, saj

nisem našla velikih razlik v odgovorih mlajših in starejših udeležencev. Sicer sem zaznala vpliv predznanja o prikazani tematiki, a ne dovolj, da bi lahko potrdila tretjo hipotezo.

Vizualizacija masovnih podatkov zagotovo pripomore k njihovemu razumevanju, saj je vizualizacijo še vedno lažje razumeti kot pa ogromne baze zapletenih in nestrukturiranih podatkov. Pa vendar, kot sem lahko ugotovila tudi v empiričnemu delu, ni samoumevno, da bo vizualizacija na tako lahek način prikazovala glavno sporočilnost. Vizualizacija ne sme biti niti preprosta, saj je potem prisotne preveč proste domišljije, niti ne sme vključevati preveč elementov, saj lahko ti odvrnejo fokus. S tem ugotavljam, da je razmerje med podatki in črnilom, ki ga je razvil Tufte, zelo pomembno načelo dobre vizualizacije. Menim, da je zelo dobro tudi, da vizualizacija vključuje kratek opis ključnih kritičnih elementov, ki bi si jih končni uporabniki vizualizacije lahko napačno interpretirali.

Sklepam lahko, da se morajo tehnike vizualizacij masovnih podatkov še nadalje razvijati predvsem v smeri učinkovitejšega prikazovanja bistvene sporočilnosti. Učinkovitost je namreč eno izmed načel dobre vizualizacije, a menim, da se da obstoječa orodja in tehnike še izboljšati, da se približajo temu načelu. Menim sicer, da razvoj tehnik in metod za vizualizacijo masovnih podatkov ne sme iti preveč v ekstrem in se dogajati prehitro, saj – kot sem povzela v teoretičnem delu – tudi preveč in prevelike novosti ne koristijo.

Kot priporočilo za nadaljnji razvoj tehnik in metod vizualizacij masovnih podatkov predlagam predvsem še večje osredotočenje na tri dejavnike, ki jih je omenjal že Wills (2012): predmet prikaza, namen vizualizacije in kdo je uporabnik vizualizacije. Na osnovi raziskave se je izkazalo za izjemenega pomena tudi, da imamo vseskozi v mislih, komu je sploh namenjena vizualizacija. Ljudje z različnim predznanjem o neki tematiki oziroma z zanimanjem za neko tematiko bodo drugače brali in zaznavali prikazano. Predlagam tudi, da ustvarjalci vizualizacij posvečajo še več pozornosti človekovemu dojemanju vizualizacij in da resnično še bolj pazijo, da vizualizacijo prikažejo na kar se da razumljiv način in da ne pretiravajo z inovativnimi elementi, ki so le videti posebno, ne povedo pa ničesar bistvenega.

LITERATURA IN VIRI

1. Albright, S. C., Winston, W. L., & Zappe, C. J. (2009). *Data Analysis & Decision Making with Microsoft® Excel* (revised 3rd ed.). Mason: South – Western Cengage Learning.
2. Ali, A., & Qadir, J. (2016). Big data for development: applications and techniques. *Big Data Analytics, 1*(2), 1-24.
3. Amster, A. (2016, 14. april). Is Your Big Data Initiative Scalable?. Najdeno 21. marca 2017 na spletnem naslovu https://www.qubole.com/blog/big-data/big-data-scalable/
4. Anagnostopoulos, I., Zeadally, S., & Exposito, E. (2016). Handling big data: research challenges and future directions. *The Journal of Supercomputing*, *72*(4), 1494-1516.
5. Andersen, R. S., Vuori, J. A., & Guillaume, X. (2015). Chromatology of security: Introducing colours to visual security studies. *Security Dialogue*, *46*(5), 440-457.
6. Ask, U., Magnusson, J., & Bredmar, K. (2016). Big Data Use in Performance Measurement and Management: A Call for Action. *Journal of Business and Economic*, *7*(3), 402-417.
7. Becker, T. (2016). Big Data Usage. V J. M. Cavanillas, E. Curry & W. Wahlster (ur.), *New Horizons for a Data-Driven Economy*, *A Roadmap for Usage and Exploitation of Big Data in Europe* (str. 143-165). Cham: Springer.
8. Belmonte, N. (2013, 16. december). Visualised: the world responds to Typhoon Haiyan on Twitter. Najdeno 4. aprila 2016 na spletnem naslovu http://twitter.github.io/interactive/philippines/
9. Berry, M. D. (2011). The computational turn: thinking about the digital humanities. *Culture Machine. 12*, 1-22.
10. Bertin, J. (1983). *Seminology of Graphics: Diagrams, Networks, Maps.* Madison: University of Wisconsin.
11. Bremer, N. (2014). Switching behavior between phone brands of the Dutch. Najdeno 25. aprila 2017 na spletnem naslovu http://bl.ocks.org/nbremer/raw/75c76f4be60fce435aba/
12. Bregar, L., Ograjenšek, I., & Bavdaž, M. (2005). *Metode raziskovalnega dela za ekonomiste: Izbrane teme.* Ljubljana: Ekonomska fakulteta.
13. Bregar, L., Pfajfar, L., & Ograjenšek, I. (2009). *Osnove statistike za ekonomiste. 1. zvezek*. Ljubljana: Ekonomska fakulteta.
14. Brenner, M. (2016, 10. oktober). 6 Data Visualization Trends for 2017. *Visual Matters.* Najdeno 28. marca 2017 na spletnem naslovu https://visualmatters.com/data-visualization-trends-2017/
15. Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of ‘big data’? *McKinsey Quarterly Report*. Najdeno 8. marca 2017 na spletnem naslovu http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/are-you-ready-for-the-era-of-big-data

16. Brown, N. (2015, 21. september). The past and future of data visualization. *Medium*. Najdeno 24. marca 2017 na spletnem naslovu https://medium.com/@uptownnick brown/the-past-and-future-of-data-visualization-d58a4b0af6b0#.xxcq86f51
17. Chaudhari, P., & Patel, B. (2017). Future of Big Data. *International Research Journal of Engineering and Technology, 4*(1), 595-597.
18. Chen, C. (2006). *Information Visualization: Beyond the Horizon* (2nd ed.). London: Springer.
19. Chen, C. L. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences, 275*, 314-347.
20. Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, *19*(2), 171-209.
21. Comai, A. (2015). Decision-Making Support: The Role of Data Visualization in Analyzing Complex Systems. *World Future Review*, *6*(4), 477-484.
22. Dean, J. (2014). *Big data, data mining and machine learning: value creation for business leaders and practitioners*. New York: Wiley.
23. Desouza, K. C. (2014). Realizing the Promise of Big Data: Implementing Big Data Projects. Najdeno 20. januarja 2016 na spletnem naslovu http://www.businessofgovernment.org/sites/default/files/Realizing%20the%20Promise%20of%20Big%20Data.pdf
24. Dur, B. I. U. (2012). Analysis of Data Visualizations in Daily Newspapers in Terms of Graphic Design. *Procedia – Social and Behavioral Sciences. 51*, 278-283.
25. Eden, B. (2005). Information Visualization. *Library Technology Reports*, *41*(1), 7-17.
26. Febretti, A., Nishimoto, A., Thigpen, T., Talandis, J., Long, L., Pirtle, J., Peterka, T., Verlo, A., Brown, M., Plepys, D., Sandin, D., Renambot, L., Johnson, A., Leigh, J. (2013). CAVE2: a hybrid reality environment for immersive simulation and information analysis. *Volume 8649. The Engineering Reality of Virtual Reality 2013.* San Francisco: Proceedings of IS&T / SPIE Electronic Imaging Science and Technology.
27. Few, S. (2009). *Data Visualization past, present and future.* b.k: IBM Cognos Innovation Center for Performance Management.
28. Friendly, M. (2008). A Brief History of Data Visualization. V C. Chen, W. Härdle & A. Unwin (ur.), *Handbook of Data Visualization* (str. 15-56). Berlin: Springer.
29. Friendly, M., Valero-Mora, P., & Ulargui, J. I. (2010). The First (Known) Statistical Graph: Michael Florent van Langren and the "Secret" of Longitude. *The American Statistician*, *64*(2), 185-191.
30. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144.
31. *Gartner.* (b.l.). Najdeno 23. junija 2015 na spletnem naslovu http://www.gartner.com/it-glossary/big-data
32. Goasduff, L. (2015, 7. oktober). Gartner Says, By 2018, Half of Business Ethics Violations Will Occur Through Improper Use of Big Data Analytics. *Gartner*

Newsroom. Najdeno 21. marca 2017 na spletnem naslovu http://www.gartner.com/newsroom/id/3144217

33. Görg, C., Pohl, M., Qeli E., & Xu, K. (2007). Visual Representations. V A. Kerren, A. Ebert, J. Meyer (ur.), *Human-Centered Visualization Environments. Lecture Notes in Computer Science* (str. 163-230). Berlin: Springer.
34. Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems, 47*, 98−115.
35. Heo, M., & Hirtle, S. C. (2001). An empirical comparison of visualization tools to assist information retrieval on the web. *Journal of the American Society for Information Science and Technology*, *52*(8), 666–675.
36. Hu, H., Wen, Y., Chua, T. S., & Li, A. X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. *IEEE Access, 2*, 652-687.
37. Jayabharathi, K. (2015). Big Data Security Challenges. *International Journal for Research in ScienceEngineering and Technology, 2*(11), 8-11.
38. Jeffery, C., & Al-Gharaibeh, J. (2015). *Writing Virtual Environments for Software Visualization.* New York: Springer.
39. Keim, D. A., Hao, M. C., Dayal, U., & Hsu, M. (2002). Pixel bar charts: a visualization technique for very large multi-attribute data sets. *Information Visualization*, *1*(1), 20-34.
40. Kennedy, H., & Hill, R. L. (2016). The Pleasure and Pain of Visualizing Data in Times of Data Power. *Television & New Media.*
41. Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-Data Applications in the Government Sector. *Communications of the ACM*, *57*(3), 78-85.
42. Kirk, A. (2016). *Data Visualization: A Handbook for Data Driven Design.* London: Sage.
43. Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers, 19*(2), 1-17.
44. Kosara, R., & Mackinlay, J. (2013). Storytelling: The Next Step for Visualization. *Computer*, *46*(5), 44-50.
45. Kosaraju, R. (2015, 15. julij). *Now or Never.* Najdeno 23. februarja 2017 na spletnem naslovu https://www.linkedin.com/pulse/because-change-things-raj-kosaraju
46. Krystian, M. (2017, 9. marec). Top Data Visualization Trends in 2017. *Infogr-am.* Najdeno 3. aprila 2017 na spletnem naslovu https://infogr.am/blog/data-visualization-trends-2017/
47. Kulyk, O., Kosara, R., Urquiza, J., & Wassink, I. (2007). Human-Centered Aspects. V A. Kerren, A. Ebert, J. Meyer (ur.), *Human-Centered Visualization Environments. Lecture Notes in Computer Science* (str. 13-76). Berlin: Springer.
48. Kung, S. Y. (2017). Discriminant component analysis for privacy protection and visualization of big data. *Multimedia Tools and Applications*, *76*(3), 3999–4034.
49. Lokhande, S., & Khare, N. (2015). An Outlook on Big Data and Big Data Analytics. *International Journal of Computer Applications*, *124*(11), 37-41.

50. Lundblad, P., & Jern, M. (2012). Visual Storytelling in Education Applied to Spatial-Temporal Multivariate Statistics Data. V J. Dill, R. Earnshaw, D. Kasik, J. Vince, P.C. Wong (ur.), *Expanding the Frontiers of Visual Analytics and Visualization* (str. 175-193). London: Springer.
51. Mack, N., Woodsong, C., MacQueen, K. M., Guest, G., & Namey, E. (2005). *Qualitative Research Methods: A Data collector's field guide.* USA: Family Health International.
52. Mangelsdorf, J. (2012). Supercomputing the Climate NASA' s Big Data Mission. *CSC World*. Najdeno 20. marca 2017 na spletnem naslovu http://assets1.csc.com/cscworld/downloads/CSCWorld_Spring2012.pdf
53. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. *Report of the McKinsey Global Institute, McKinsey & Company.* Najdeno 10. maja 2015 na spletnem naslovu http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
54. Marconi, K., Dobra, M., & Thompson, C. (2013). The Use of Big Data in Healthcare. V J. Liebowitz (ur.), *Big Data and Business Analytics* (str. 229-248). Boca Raton: Taylor & Francis Group.
55. Marr, B. (2014, 6. marec). Big Data: The 5 Vs Everyone Must Know. Najdeno 17. decembra 2016 na spletnem naslovu https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know
56. Marr, B. (2015, 7. maj). From Farming To Big Data: The Amazing Story of John Deere. *Data Science Central*. Najdeno 20. marca 2017 na spletnem naslovu http://www.datasciencecentral.com/profiles/blogs/from-farming-to-big-data-the-amazing-story-of-john-deere
57. Marr, B. (2016a, 15. marec). 17 Predictions About The Future Of Big Data Everyone Should Read. *Forbes*. Najdeno 21. marca 2017 na spletnem naslovu https://www.forbes.com/sites/bernardmarr/2016/03/15/17-predictions-about-the-future-of-big-data-everyone-should-read/#18c61ac61a32
58. Marr, B. (2016b, 4. maj). How VR Will Revolutionize Big Data Visualizations. *Forbes.* Najdeno 3. aprila 2017 na spletnem naslovu https://www.forbes.com/sites/bernardmarr/2016/05/04/how-vr-will-revolutionize-big-data-visualizations/2/#776c167211f4
59. McCosker, A., & Wilken, R. (2014). Rethinking 'big data' as visual knowledge: the sublime and the diagrammatic in data visualisation. *Visual Studies, 29*(2), 155-164.
60. Milakovich, M. E. (2012). Anticipatory Government: Integrating Big Data for Smaller Government.. *Internet, Politics, Policy 2012: Big Data, Big Challenges*. Najdeno 24. februarja 2017 na spletnem naslovu http://ipp.oii.ox.ac.uk/sites/ipp/files/documents/Anticipatory%20Government%20Oxford.pdf

61. Miller, J., & Glassner, B. (2011). The »Inside« and the »Outside«: Finding Realities in Interviews. V D. Silverman (ur.), *Qualitative Research* (3rd ed.) (str.131-148). London: Sage.
62. Munné, R. (2016). Big Data in the Public Sector. V J. M. Cavanillas, E. Curry & W. Wahlster (ur.), *New Horizons for a Data-Driven Economy*, *A Roadmap for Usage and Exploitation of Big Data in Europe* (str. 195-208). Cham: Springer.
63. Olshannikova, E., Ometov, A., Koucheryavy, & Y., Olsson, T. (2015*).* Visualizing Big Data with augmented and virtual reality: challenges and research agenda. *Journal of Big Data*, *2*(22), 1-27.
64. Partl, V., Gratzl, S., Streit, M., Wassermann, A. M., Pfister, H., Schmalstieg, D., & Lex, A. (2016). Pathfinder: Visual Analysis of Paths in Graphs. *Computer Graphics Forum*, *35*(3), 71-80.
65. Porambge, P., Ylianttila, M., Schmitt, C., Kumar, P., Gurtov, A., & Vasilakos, A. V. (2016). The Quest for Privacy in the Internet of Things. *IEEE Cloud Computing*, *3*(2), 36-45.
66. Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics, 32*(4), 240-241.
67. Rahm, E., & Do, H. H. (2000). Data cleaning: problems and current approaches. *IEEE Data Engineering Bulletin*, *23*(4), 3-13.
68. Rist, O. (2016, 15. avgust). The Best Data Visualization Tools of 2016. *PC*. Najdeno 24. marca 2017 na spletnem naslovu http://www.pcmag.com/roundup/346417/the-best-data-visualization-tools
69. Rodríguez-Mazahua, L., Rodríguez-Enríquez, C.A., Sánchez-Cervantes, J. L., Cervantes, J., García-Alcaraz, J. L., & Alor-Hernández1, G. (2016). A general perspective of Big Data: applications, tools, challenges and trends. *The Journal of Supercomputing*, *72*(8), 3073-3113.
70. Seeger, M. (2009). *Key-Value stores: a practical overview*. Stuttgart: Computer Science and Media.
71. Segel, E., & Heer, J. (2010). Narrative Visualization: Telling Stories with Data. *Visualization and Computer Graphics*, *16*(6), 1139-1148.
72. Sheny, Z., & Maz, K. L. (2007). Path visualization for adjacency matrices. *Proceeding EUROVIS'07 Proceedings of the 9th Joint Eurographics / IEEE VGTC conference on Visualization* (str. 83-90). Norrköping: EUROGRAPHICS The European Association for Computer Graphics.
73. Telea, A. C. (2014). *Data Visulization. Principles and practice* (2nd ed.). Boca Raton: Taylor & Francis Group.
74. Towler, W. (2015, januar/febraur). Data Visualization: The future of data visualization. *Analytics*. Najdeno 24. marca 2017 na spletnem naslovu http://analytics-magazine.org/data-visualization-the-future-of-data-visualization/
75. Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.
76. Tufte, E. R. (1990). *Envisioning Information.* USA: Graphics Press.

77. Ularu, E. G., Puican, F. C., Apostu, A., & Velicanu, M. (2012). Perspectives on Big Data and Big Data Analytics. *Database Systems Journal*, *3*(4), 3-14.
78. Vogrinc, J. (2008). *Kvalitativno raziskovanje na pedagoškem področju*. Ljubljana: Pedagoška fakulteta.
79. Walker, M. (2013, 13. avgust). Batch vs. Real Time Data Processing. *Data Science Central.* Najdeno 21. marca 2017 na spletnem naslovu http://www.datasciencecentral.com/profiles/blogs/batch-vs-real-time-data-processing
80. Wang, R. (2012, 27. februar). Monday's Musings: Beyond The Three V's of Big Data—Viscosity and Virality. *Softwareinsider.org.* Najdeno 7. avgusta 2015 na spletnem naslovu http://blog.softwareinsider.org/2012/02/27/mondays-musings-beyond-the-three-vs-of-big-data-viscosity-and-virality/
81. Warden, P. (2011). *Big Data Glossary*. USA: O'Reilly.
82. Ware, C. (2013). *Information Visualization: Perception for Design* (3rd ed.). San Francisco: Morgan Kaufmann.
83. Wiener, N. (1964). *I am Mathematician.* Cambridge: MIT Press.
84. Wills, G. (2012). *Visualizing Time: Designing Graphical Representations for Statistical Data*. New York: Springer.
85. Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques* (2nd ed.). San Francisco: Morgan Kaufmann.
86. Wojtkowski, W., & Wojtkowski, W. G. (2002). Storytelling: its role in information visualization. Najdeno 28. marca 2017 na spletnem naslovu http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.4771&rep=rep1&type=pdf
87. Wortmann, F., & Flüchter, K. (2015). Internet of Things. Technology and Value Added. *Business & Information Systems Engineering, 57*(3), 221-224.
88. Xu, Z., & Shi, Y. (2015). Exploring Big Data Analysis: Fundamental Scientific Problems. *Annals of Data Science*, *2*(4), 363-372.
89. Zhong, R.Y., Lan, S., Xu, C., Dai, Q., & Huang, G. Q. (2016). Visualization of RFID-enabled shopfloor logistics Big Data in Cloud Manufacturing. *The International Journal of Advanced Manufacturing Technology, 84*(1), 5-16.
90. Zhu, Y. (2007). Measuring Effective Data Visualization. V G.Bebis et al. (ur), *Advances in Visual Computing*, *Lecture Notes in Computer Science, vol. 4842* (str. 652-661). Berlin: Springer.
91. Žumer, M., & Merčun, T. (2008). Vizualizacija informacij v sistemih za poizvedovanje. *Zveza bibliotekarskih društev Slovenije, 52*(2/3), 95-114.

PRILOGE

KAZALO PRILOG

PRILOGA 1: Vprašalnik vizualizacijske skupine 1

PRILOGA 2: Vprašalnik skupine s ključnimi besedami 4

PRILOGA 1: Vprašalnik vizualizacijske skupine

Pozdravljeni. Najlepša hvala, ker ste si vzeli čas, da sodelujete v empiričnem delu moje magistrske naloge na temo vizualizacije masovnih podatkov. V njem želim preučiti, do kakšne mere so vizualizacije masovnih podatkov razumljive in kako jasno podajajo svojo glavno temo prikaza, zato se vam že vnaprej zahvaljujem, da mi boste pomagali pri tem delu.

Pogovor bom tudi snemala, da bom podatke lahko kasneje analizirala za raziskovalni namen. Vaši odgovori so zaupni ter bodo uporabljeni samo za raziskovalni namen mojega magistrskega dela. Prosim vas, da le odgovarjate iskreno po svojih najboljših močeh. Vodila vas bom skozi več nalog.

Primer 1: Aktivnost na tviterju po katastrofalnem tajfunu na Filipinih novembra 2013

V okviru prve naloge vam bom najprej opisala resnični dogodek, vi si ga poskusite čim bolje zamisliti, ker vas bom prosila, da ga boste potem poskusili tudi grafično prikazali na listu papirja. Opis resničnega dogodka vam bom predrožila tudi v pisni obliki v primeru, da boste morali še enkrat pogledati opis.

Za to nalogo so predvidene tri minute, ki se pričnejo šteti zdaj. V primeru vprašanj, boste seveda imeli potem na voljo manj časa za vaš grafični prikaz.

Filipine je novembra 2013 prizadel katastrofalen tajfun. Pomoč in izrazi sožalja ter sučustvovanja so prihajali od vsepovsod, med drugimi je to izrazilo tudi na tisoče ljudi preko tviterja. To so storili preko tako imenovanih tvitov - kratkih sporočil, kjer so v različnih jezikih iz vseh koncev sveta, pošiljali svoje izraze podpore in sočustvovanja. Radi bi grafično prikazali, pošiljanje teh tvitov na Filipine iz vseh kontinentov. Prosim vas torej, da na list papirja skicirati omenjene tvite.

Najlepša hvala za vašo vizualizacijo, tukaj pred vami pa je sedaj vizualizacija ki jo je objavil Nicolas Belmonte na internetu po katastrofalnem dogodku, seveda pa je zanjo namenil kar precej časa.

Prosim vas če pogledava, kaj lahko najdeva skupnega med vašo in sedaj predstavljeno vizualizacijo. Poizkusiva najti vsaj tri podobnosti in tri razlike.

Kaj pa se vam zdi boljše pri vaši in kaj pri dejanski vizualizaciji?

Kaj se vam zdi najtežje in najmanj razumljivo pri dejanski vizualizaciji? Bi ugotovili, kaj prikazuje vizualizacija, če ne bi poznali predzgodbe – zakaj?

Primer 2: Prehajanje med blagovnimi znamkami mobilnih telefonov med nizozemskimi uporabniki

Prosila vas bom, da naštevate vse kar vam pade na misel ob tej vizualizaciji. Na voljo imate pet minut, besede, ki jih boste naštevali, bom sama beležila sproti.

Sedaj pa teh ključne besede poizkusite povezati skupaj in oblikovati kaj bi lahko bil naslov takšne vizualizacije.

Primer 3: Prikaz logističnih poti s pomočjo brezžičnih radijskih valov

Prosim, če si lahko po svojih najboljših močeh zamišljate naslednjo situacijo in poizkusite to na list papirja tudi vizualizirati. Opis situacije vam podajam tudi na tem listu papirja, da ga boste lahko še enkrat prebrali, v kolikor boste želeli. Na voljo za vizualizacijo boste imeli 5 minut, ki se pričnejo upoštevati zdaj. V primeru vprašanj bodo se le ta upoštevala v 5 razpoložljivih minut.

Podjetje Huaiji iz Kitajske je specializirano v proizvodnji ventilov motorja, pri čemer uporablja najnovejše tehnologije. Od leta 2003 naprej so svoje napore usmerili tudi v izboljšanje notranjih procesov. Da bi pridobili podatke o trenutnem stanju in jih uporabili tudi za izboljšanje, so zgradili tehnologije brezžičnih radijskih valov. V primeru logističnih operacij v enem proizvodnem oddelku to pomeni, da so na različne korake proizvodnje in premikanja produkta iz ene v drugo fazo proizvodnje namestili brezžične vmesnike. Ti zaznane informacije direktno pošiljajo v bazo, kjer se le ti zbirajo.

Prosim vas, da na list papirja vizualizirate naslednje.V proizvodnem obratu je sedem različnih faz, preko katerih lahko gredo izdelki. Vsaka izmed sedmih proizvodnih faz ima še podfaze, v katerih potekajo različna dela proizvodnje izdelka. Število podfaz v različnih fazah je različno:
faza 1: 8 podfaz
faza 2: 6 podfaz
faza 3: 6 podfaz
faza 4: 4 podfaz
faza 5: 9 podfaz
faza 6: 10 podfaz
faza 7: 4 podafaz

Zamislite si proizvodnjo enega paketa izdelkov. Želimo prikazati, kako se ta paket izdelkov tekom proizvodnje premika med različni fazami in podfazi. Predstavljajmo si, da gre ta specifični paket izdelkov tekom proizvodnje preko 18 podfaz, zato se mora 18-krat premakniti, da na koncu dobimo končni paket izdelkov.

Najlepša hvala za vašo vizualizacijo, tukaj pred vami pa je sedaj vizualizacija ki je bila objavljena v strokovnem članku na zgoraj opisano tematiko, seveda so zanjo porabili nekoliko več časa in napredno tehnologijo.

Prosim vas če pogledava, kaj lahko najdeva skupnega med vašo in sedaj predstavljeno vizualizacijo. Poizkusiva najti vsaj tri podobnosti in tri razlike.

Kaj pa se vam zdi boljše pri vaši in kaj pri dejanski vizualizaciji?

Kaj se vam zdi najtežje in najmanj razumljivo pri dejanski vizualizaciji? Bi ugotovili, kaj prikazuje vizualizacija, če ne bi poznali predzgodbe – zakaj?

Konec: Splošno

Sedaj ko sem z vami šla skozi tri primere vizualizacij masovnih podatkov, bi vas prosila tudi, če mi lahko prosim poveste še vaše mnenje, kateri izmed primerov vam je bil najbolj razumljiv iz vidika, kako hitro vam je bilo jasno kaj prikazuje vizualizacija, in kateri najmanj razumljiv. Zakaj?

Najlepša hvala še enkrat za vaše sodelovanje.

PRILOGA 2: Vprašalnik skupine s ključnimi besedami

Pozdravljeni. Najlepša hvala, ker ste si vzeli čas, da sodelujete v empiričnem delu moje magistrske naloge na temo vizualizacije masovnih podatkov. V njem želim preučiti, do kakšne mere so vizualizacije masovnih podatkov razumljive in kako jasno podajajo svojo glavno temo prikaza, zato se vam že vnaprej zahvaljujem, da mi boste pomagali pri tem delu.

Pogovor bom tudi snemala, da bom podatke lahko kasneje analizirala za raziskovalni namen. Vaši odgovori so zaupni ter bodo uporabljeni samo za raziskovalni namen mojega magistrskega dela. Prosim vas, da le odgovarjate iskreno po svojih najboljših močeh. Vodila vas bom skozi več nalog.

Primer 1: Aktivnost na tviterju po katastrofalnem tajfunu na Filipinih novembra 2013

Pri tej nalogi vas bom prosila, da boste naštevali kar vam bo padlo na misel, ob naslednji vizualizaciji, ki vam jo bom pokazala. Na voljo imate pet minut, besede, ki jih boste naštevali, bom sama beležila sproti. Čas se pričnete šteti sedaj.

Sedaj pa te ključne besede poizkusite povezati skupaj in oblikovati kaj bi lahko bil naslov takšne vizualizacije. Če vam pomagam z naslednjimi podatki, da ima vizualizacija opravka v z nekim naravnim pojavom novembra 2013 v Jugovzhodni Aziji in da prikazuje tudi nekaj v povezavi s socialnimi omrežji. Na voljo imate največ pet minut.

Primer 2: Prehajanje med blagovnimi znamkami mobilnih telefonov med nizozemskimi uporabniki

Opisala vam bom anketo, ki je bila leta 2014 izvedena na Nizozemskem, med prebivalci, ki uporabljajo mobilne telefone. Prosim pozorno poslušajte besedilo in kaj vam bom dala kot navodilo, da grafično prikažete na listu papirja. Opis ankete in navodila za grafični prikaz vam bom predložila tudi na posebnem listu papirja, da boste lahko vmes še pogledali nanj.

Za to nalogo je predvidenih osem minut, ki se pričnejo šteti zdaj. V primeru vprašanj, boste seveda imeli potem na voljo manj časa za vaš grafični prikaz.

Skoraj vsak Nizozemec uporablja mobilni telefon in veliko uporabnikov zamenja telefon takoj, ko jim to omogoča pogodba. Ampak, ali ljudje preidejo tudi na drugo blagovno znamko in kako se ta prehod razlikuje med različnimi znamkami?

Na to temo so izvedli anketo, potem pa so analizirali odgovore in povezave med anketirančevo trenutno in prejšnjo znamko mobilnega telefona.

Grafično prikažite, da se trenutno uporablja sedem večjih blagovnih znamk, vse ostale pa lahko združimo v eno skupino: Nokia, Huawei, HTC, Apple, LG, Samsung, Sony in ostalo. Trenutno se tržni deleži sledijo v naslednjem vrstnem redu (od največjega do najmanjšega): Samsung, Apple, Nokia, HTC, LG in Sony.

Prikažite tudi, da je kar nekaj anketirancev preklopilo iz Nokie na Samsung, ter iz Samsunga na Nokio. Delež tistih, ki so preklopili iz Nokie na Samsung, je večje, kot v obratni smeri.

Nokia je izgubila veliko uporabnikov, ki so prešli na Samsung, nekaj jih je prešlo na HTC, LG in Huawei. Nekaj pa so jih tudi obdržali.

Huawei je nova znamka, ki je vstopila na trg, nanj so preklopili največ uporabniki Samsunga, nihče od uporabnikov Appla ni naredil tega preskoka.

HTC je pridobil nekaj Nokinih uporabnikov in LG uporabnikov, nekaj jih je obdržal vedno pa ima celotno majhen delež na celotnem trgu. še

Apple ni izgubil skoraj nobenega uporabnika, jih je pa kar nekaj pridobil od Samsunga, Nokie in ostalih.

LG ima mešanico. Nekaj uporabnikov je prešlo na Samsung, Apple, HTC in Huawei, medtem ko jih je nekaj pridobil od Nokie, Sonya in ostalih.

Samsung je največ novih kupcev pridobil od Nokie in ostalih manjših blagovnih znamk, pridobil pa je še od vseh ostalih večjih znamk razen Appla in Huawei.

Sonny je izgubil nekaj kupcev, ki so odšli k Samsungu, Noki, Applu, LG in Huawei.

Najlepša hvala za vašo vizualizacijo, tukaj pred vami pa je sedaj vizualizacija ki jo je objavil Nadieh Bremer, seveda pa je vanjo vložil kar precej časa.

Prosim vas če pogledava, kaj lahko najdeva skupnega med vašo in sedaj predstavljeno vizualizacijo. Poizkusiva najti vsaj tri podobnosti in tri razlike.

Kaj pa se vam zdi boljše pri vaši in kaj pri dejanski vizualizaciji?

Kaj se vam zdi najtežje in najmanj razumljivo pri dejanski vizualizaciji? Bi ugotovili, kaj prikazuje vizualizacija, če ne bi poznali predzgodbe – zakaj?

Primer 3: Prikaz logističnih poti s pomočjo brezžičnih radijskih valov

Izvolite, tukaj pred vami je vizualni prikaz. Prosila vas bom, da naštevate vse kar vam pade na misel ob tej vizualizaciji. Na voljo imate pet minut, besede, ki jih boste naštevali, bom sama beležila sproti. Prosim, navedite tudi naslov vizualizacije, po vašem mnenju.

Konec: Splošno

Sedaj ko sem z vami šla skozi tri primere vizualizacij masovnih podatkov, bi vas prosila tudi, če mi lahko prosim poveste še vaše mnenje, kateri izmed primerov vam je bil najbolj razumljiv iz vidika, kako hitro vam je bilo jasno kaj prikazuje vizualizacija, in kateri najmanj razumljiv. Zakaj?

Najlepša hvala še enkrat za vaše sodelovanje.